5-13-2011

Caspase-7 Loop Conformations as a Means of Allosteric Control

Witold Andrej Witkowski
University of Massachusetts Amherst, wwitkows@chem.umass.edu

CASPASE-7 LOOP CONFORMATIONS AS A MEANS OF ALLOSTERIC CONTROL

A Dissertation Presented

by

WITOLD A. WITKOWSKI

Submitted to the Graduate School of the
University of Massachusetts, Amherst in partial fulfillment
of the requirements for the degree of

DOCTOR OF PHILOSOPHY

May 2011

Chemistry Department

CASPASE-7 LOOP CONFORMATIONS AS A MEANS OF ALLOSTERIC CONTROL

A Dissertation Presented

by

WITOLD A. WITKOWSKI

Approved as to style and content by:

_____
Jeanne A. Hardy, Chair

_____
Scott C. Garman, Member

_____
Lynmarie K. Thompson, Member

_____
Ramgopal R. Mettu, Member

_____
Craig T. Martin, Department Head
Chemistry Department

DEDICATION

To my Dziadziu Heniu, who taught me to never let myself be average, and always said that no matter how hard you are trying, try harder.

ACKNOWLEDGEMENTS

"If we knew what it was we were doing, it would not be called research, would it?"
-Albert Einstein

Research does not happen in a vacuum, and so I know that I could not have done anything without the help, support, and encouragement from Jeanne and my Hardy Lab family. Through all the fun times when data seemed to be coming easily, and through all the more difficult times when things just wouldn't work, I knew that I could always talk these things through. I enjoyed being the 'big kids' with Sravanti and Kristen, trying to look out for the newer graduate students, thinking of ways to make things more fun, and recruiting out butts off to grow the lab. I'll never forget learning everything from scratch together, becoming gel running mavens in only 6 months ☺.

To my UMass Family. Dr. Scott Garman, who's guidance and ideas often helped me through some tricky parts of my doctoral work. I cannot forget my stellar committee members Dr. Lynmarie Thompson and Dr. Ramgopal Mettu, who knew how to push me to excel, and helped me focus my research. I'll never forget during my ORP, when Scott asked me a question I was totally unprepared for. At that moment, I was convinced I had failed the exam, but when I came back into the room, and was told that I had in fact, *passed*, I had to sit down, and Lynmarie handed me a cookie, and Jeanne handed me water, both afraid I would faint.

To my teachers. Dr. Howard Stidham, who with kind words and brilliant intellect convinced me that I should attend the University of Massachusetts. Dr. Gamil Guirgis, because without his great advice, encouragement and help I would have never thought to join graduate school. My spectacular 7th grade science teacher, Mr. Michael Beitzel, who stoked my curiosity about the natural world, encouraged me to experiment and learn on my own, and never thought any of my questions were too silly to answer.

To Mom, Dad, Rose, Ania, Asia, Marcin, Ciocia Jola and all my family. I could not have done this without your support, guidance, Love, and strength. I could go on for pages with examples of ways you made me feel better about myself, and about what I was doing in my life. Thank you for everything you had to do for me to be able to study. There are no words to express my gratitude. Kocham was wszystkich bardzo i dziękuję jeszcze raz.

ABSTRACT

CASPASE-7 LOOP CONFORMATIONS AS A MEANS OF ALLOSTERIC CONTROL

MAY 2011

WITOLD WITKOWSKI, B.S. COLLEGE OF CHARLESTON

Ph.D. UNIVERSITY OF MASSACHUSETTS AMHERST

Directed by: Professor Jeanne A. Hardy

The caspase family of proteins is critical to biological understanding, because they serve as the final arbiters of life and death, being the initiators and executioners of cell death. Specifically, caspase-7 plays a key role in apoptosis, however its full complement of targets within the cell has not yet been elucidated, nor has its function been targeted by drug design efforts. These factors stem from the lack of fundamental understanding of the structural dynamics of the protein, including the mobile loops that constitute the active site binding groove of caspase-7, and their ability to modulate the function of the protein. In this work, we describe the importance of the entire loop bundle for catalysis, demonstrate a novel approach for allosteric control using loop movement, develop computational methods to engineer a new binding site for an allosteric effector and discover a hereunto unseen native disulfide within caspase-7 that may contribute to specificity and catalysis. The information obtained within this study is applicable for not only the study of caspase-7, but also the greater field of apoptosis research.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER I

INTRODUCTION

The ability to dynamically and independently modulate the function of proteins for either elucidative reasons or curative treatments would change the world of medicine. Advances in structural biochemistry, combined with novel techniques will allow new problems to be tackled. In this work, I present advancements in the field of protein chemistry that develop new insights into the role of the loops in caspase-7, and their ability to control caspase-7.

Apoptosis

For many years, it was tacitly assumed that due to continuous cellular proliferation, a mechanism for programmed cell death must exist. Until cell morphologies during cell death were rigorously defined, programmed cell death was broadly described as coagulative necrosis active site[1]. The morphological changes in two types of cell death, necrosis and apoptosis, lead to the distinction in the classes. During necrosis, cells swell and burst, leaving cellular debris, which often results in an inflammatory response. Conversely, apoptosing cells shrink, and are rapidly phagocytosed in an orderly fashion. Since Kerr and coworkers coined the term apoptosis in their seminal 1972 paper, the highly conserved cellular machinery responsible for this ordered process has been broadly studied across many organisms, and its evolutionarily conserved pathways have been elucidated.

Apoptosis' importance to cellular function and organism development cannot be understated, its regulation is necessary for both death, and life in cells. Apoptosis is most associated with cellular demise, however regulated cell death is necessary for proper organism development. For instance, in mice deficient a cell death protein (caspase-3), organism death occurs early in embryonic development with an over abundance of neuroepithelial cells[2]. This suggests that cell death is necessary in neurological development. In another experiment, irreversible broad spectrum peptidic inhibitors of apoptosis were introduced into chicken embryos, and the footpads were studied[3]. In embryos treated with the inhibitors, evidence of interdigital cell death decreased, leaving the embryo with webbed feet (Figure 1). The role of apoptosis is not limited to limb development, as many have shown that regulated cell death plays a key role in sexual differentiation[4-7], and terminal cell differentiation. A erythrocyte's final developmental stage involves the loss of organelles and nucleus, but remain viable and continue to metabolize; conversely keratinocytes form layers of squamous skin layers at their final differentiation stage. These two near end of life or alternate G1 exit stage processes are thought to be the result of modified programmed cell death[8,9].

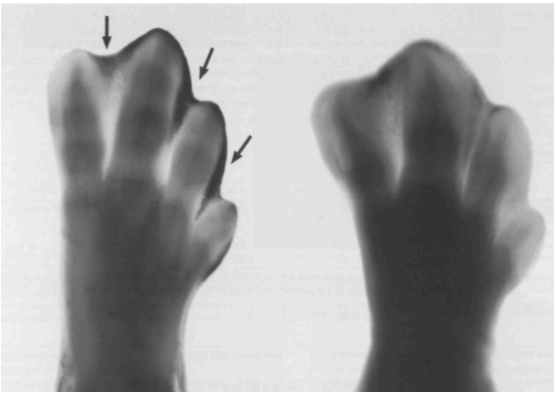


Figure 1: Inhibition of apoptosis in development.

Interdigital cell death (dark staining, arrows) is retarded in the presence of apoptotic inhibitors[3].

Disorders Associated with Apoptosis

Apoptosis's role in the life/death decision is critical, and cellular havoc occurs when apoptosis goes wrong. Errors in apoptosis have been linked to many diseases, as the delicate life/death balance is disturbed. During cell proliferation, a number of regulators and growth factors check the cell cycle to maintain order[10]. The disease most classically associated with down-regulated cell death is cancer. Though there exist many variants of malignancy, all stem from a common inability to undergo apoptosis[11]. Disorders typically associated with rapid cell death are heart attacks and stroke. Notably, these diseases are primarily caused by lack of blood flow to the region, leading the affected cells to necrose. However, tissue not directly affected by the restricted blood flow, but in the vicinity of necrotic tissue has been shown to die by apoptosis[12,13], and directly inhibiting apoptosis in these cells can limit the damage[14]. Less obvious diseases such as viral infections and their associated cellular processes have also been linked to apoptosis[15,16].

Caspases

The amazing molecular scale machinery that is responsible for executing the balance of life and death in cells are called caspases, from **C**ysteine-**Asp**artate-Prote**ases**. The origin of caspases and their role in the apoptotic pathway was first reported by Robert Horvitz and coworkers, while studying the *C. elegans* death gene *CED-3*[17]. Horvitz was subsequently awarded the Nobel Prize for the elucidation of the apoptotic pathway. Horvitz and coworkers noted a predicted protein sequence similarity between *CED-3* and

the human and murine interlukin-1β converting enzyme. This enzyme had been characterized as a heterodimeric cysteine protease[18], and later renamed caspase-1. Discovery of other caspases followed soon after, and necessitated the distinction of apoptotic caspases (-2, -3, -6, -7, -8, -9, -10) versus inflammatory caspases (-1, -4, -5, -11, -12, -13, -14) (For review, see[19-22]).

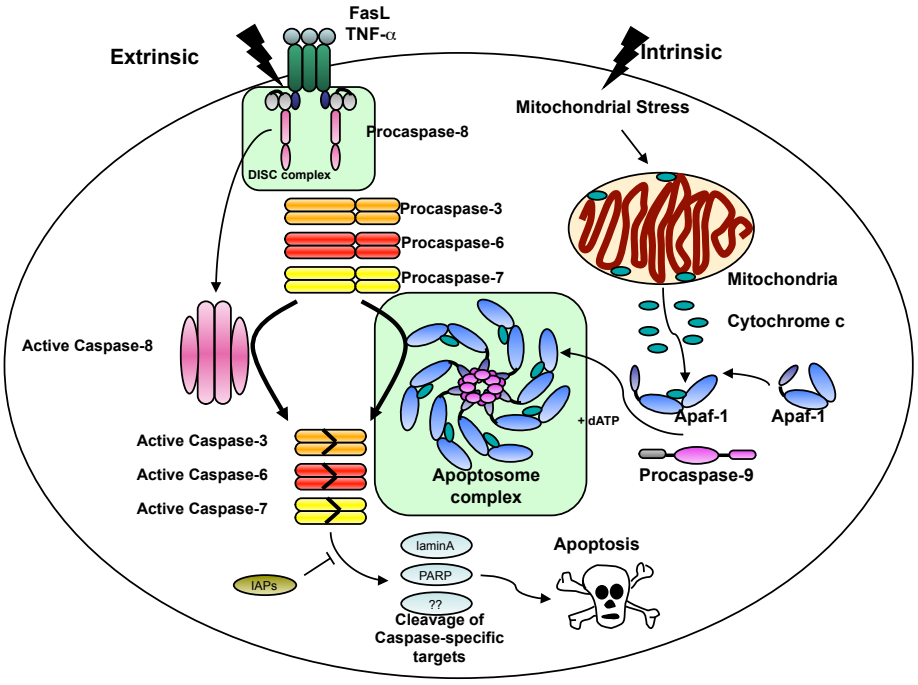


Figure 2: Differences in intrinsic versus extrinsic caspase activation

Caspase and thus apoptosis activation can occur either from extrinsic (external) or intrinsic (internal) cellular stresses. Based on the nature of the stress, initiation different pathways are followed. In the intrinsic pathway, mitochondrial stress causes the release of cytochrome c, which triggers the formation of the apoptosome complex, leading to activation of the downstream executioner caspases by caspase-9. In the extrinsic pathway, the DISC complex is activated by an extracellular signal, which then causes the maturation of caspase-8, which cleaves executioner caspases.

Apoptotic caspases can be classified as either initiator (-2, -8, -9) or executioner (-3, -6, -7). Like their necrotic cousins, apoptotic caspases are expressed in the cytosol as catalytically inactive zymogens, referred to as pro-enzymes. Caspases are expressed as a single chain polypepide that requires cleavage for maturation. Broadly described, this

single chain is composed of three sections, the N-terminal peptide, large subunit, and small subunit. During maturation and activation, the N-terminal peptide is cleaved off, as is an inter-subunit linker. The length of the N-terminal peptide and inter-subunit linker and the order of cleavage are specific to each caspase. (For review see[19])

The activation of these proteolytic enzymes is cellular stress-dependent and the apoptotic pathway which activates the caspases varies depending on the nature of the stress. In the extrinsic pathway, an extracellular signal is received by TNF-α (tumor necrosis factor), which is spatially associated to procaspase-8 in the DISC (death inducing signal complex). This signal causes the recruitment of procaspase-8 to the complex and leads to the activation of pro-caspase-8. The activation of procaspase-8 is believed to be caused by auto-processing[23] due to high local concentrations of pro-form caspase-8. Once activated, the mature caspase-8 cleaves the downstream or executioner caspases (-3, -6, -7).

In the intrinsic pathway, mitochondrial stress induces cytochrome c release, which bind to Apaf-1, forcing a conformational change in Apaf-1. This Apaf-1:cytochrome c complex facilitates the formation of the apoptosome, which is a heptamer of Apaf-1 and cytochrome c in the presence of ATP or dATP[24,25]. The apoptosome then recruits pro-caspase-9, and it is thought that the close proximity of two monomers of pro-caspase-9 induce the intersubunit cleavage activation of caspase-9. Caspase-9, like caspase-8, then cleaves the executioner caspases thus activating them. Since the executioner caspases are apoptosis pathway independent, their activation typically sounds the death knell for the cell. Because of their ability to kill cells or contribute to cell survival, the ability to control caspase function is of great utility. In this thesis, we focus on one of the

executioner caspases, caspase-7, exploring the molecular details of its function and engineering a redox sensitive version of caspase-7.

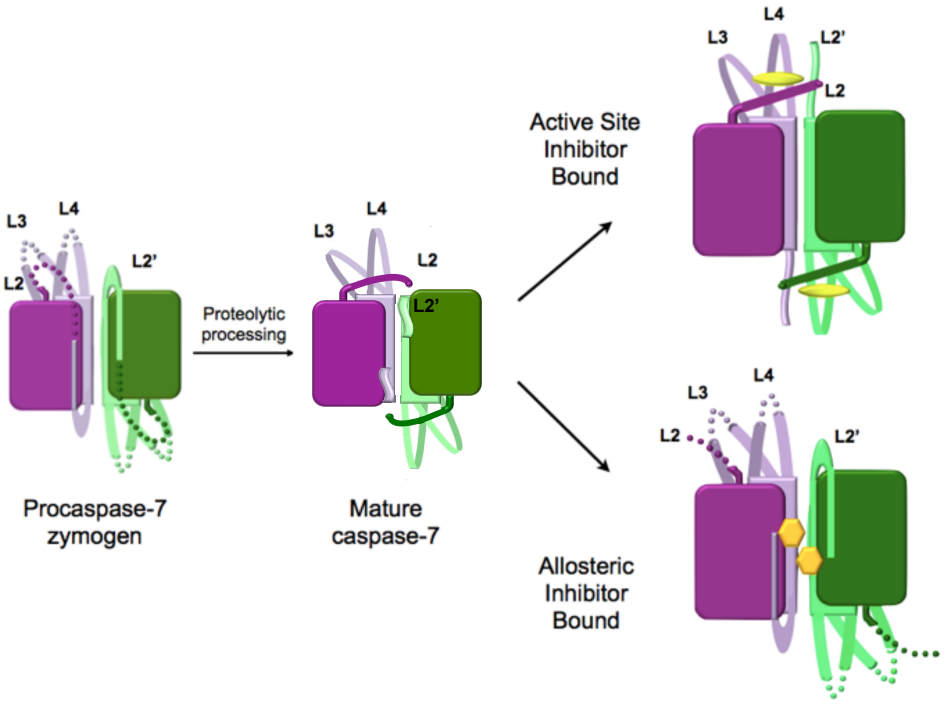


Figure 3: Caspase-7 conformations

Caspase-7 is expressed as a single chain polypeptide which dimerizes and forms the procaspase-7 **zymogen** where loops L2, L3, and L4 are formed but are in a catalytically incompetent conformation. Proteolytic processing cleaves the inter-subunit linker on L2 creating the **mature** caspase, which is formed from two small subunits, and two large subunits. Loops L2 and L2' from the opposite side of the dimer are liberated during this cleavage event, but are not in catalytically competent conformation. In the **active** form of caspase-7, binding a covalent inhibitor at the active site cysteine promotes the ordering of the active site loop bundle, where L2, L2', L3 and L4 are clearly visible in the crystal structure. Conversely, in the **allosterically inhibited** form where small molecule inhibitors DICA or FICA are bound covalently to cysteines at the dimer interface (C290), L2' adopts a conformation similar to the **zymogen** form, while L2, L3, and L4 are disordered.

Executioner caspases are obligate homodimers[26,27], each composed of a large and small subunit which are formed during cleavage activation by an upstream caspase such as caspase-8 or -9. It is notable, however, that caspases recombinantly expressed in *E. coli* self activate by an unknown mechanism. The substrate-binding groove is made up of flexible loops on opposite sides of the two-fold symmetrical dimer. The interface of the

homodimer of heterodimers creates a large cavity formed by the interface between the small subunits composed of anti-parallel β strands (Figure 4). While the crystal structures of caspase-3[28-34] and -6[35,36,27] exist, caspase-7 has more structural information available because the three-dimensional structure of multiple conformations has been solved previously by x-ray crystallography[37-42]. The notable differences between the structures lie in the arrangement of the loops that ultimately form the substrate-binding groove. In the inactive zymogen caspase, the intersubunit linker that is part of loop L2 is uncleaved, lying over the cavity at the interface of the small subunits. (Figure 3) After processing, L2 loop is split into L2 and L2'. In the crystal structure of the mature, but catalytically incompetent form of caspase-7, loops, L2, L3, and L4 are ordered, but L2' loops lie over the allosteric sites and the substrate binding grooves are not competent to accept substrate. When crystallized with a covalently bound tetrapeptide suicide substrate (DEVD-CHO), the loops appear fully ordered[39], however, the active site geometry of the cysteine-histidine dyad is not favorable for catalysis since the catalytic thiol is covalently modified by the bound substrate. Binding to the aldehyde containing substrate pushes the catalytic thiol out of position for catalysis. The active-site cysteine (Cys186 in caspase-7 numbering) lies at the base of L2, and the trajectory of the distal end of L2 causes interactions between it and L2' from the opposite side of the dimer. Caspase-7 has also been crystallographically observed to adopt an allosterically inhibited conformation when small molecules DICA or FICA are bound to cysteine 290[42] as represented in Figure 3. In this structure, L2' adopts a loop trajectory similar to that of the procaspase-7 structure or the mature apo structure, while loops L2, L3, and L4 are disordered.

These structural data provided many insights into how the caspase works, and have

facilitated study of the active site and the allosteric cavity[42].

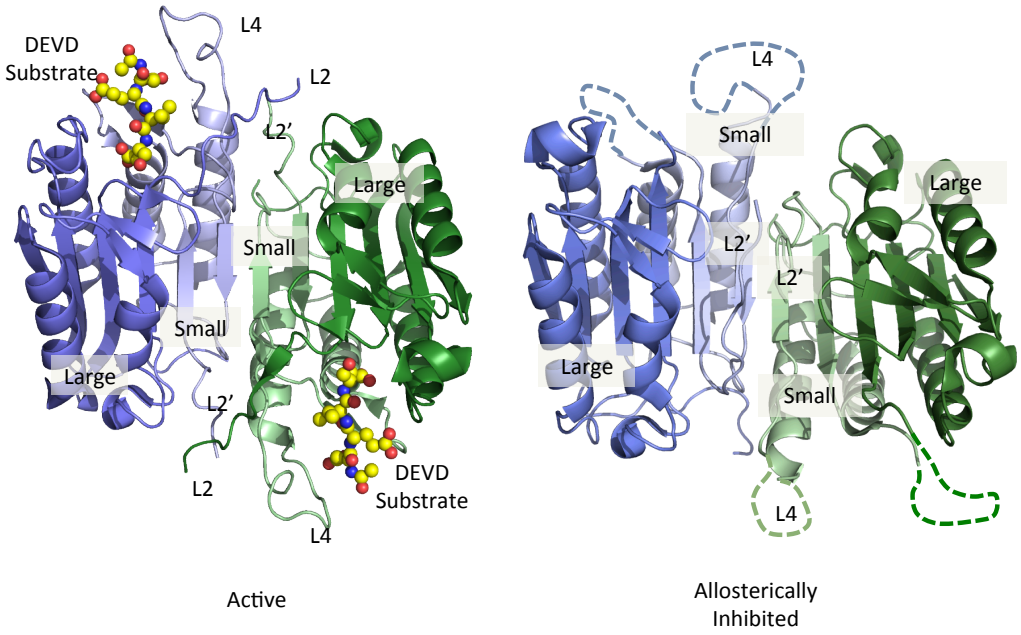


Figure 4: Caspase-7 structure

Caspase-7 structure is composed of two large subunits (darker color) and two small subunits (lighter color) forming a dimer with two fold symmetry. (A) In the active structure of caspase-7, a DEVD substrate (spheres) is covalently bound to the active site cysteine, which lies at the base of loop L2. Loop L2 and L2' along with L3 and L4 arrange to form the substrate binding groove. (B) In allosterically inhibited structure of caspase-7, L2' from both sides of the dimer lie over the dimer interface in a catalytically inactive form. Additionally, the loops of the active site binding groove are disordered (dashes).

The central cavity formed between the two halves of the homodimer can be harnessed for allosteric regulation. In the crystallographic model, allosteric inhibition can be triggered by the introduction of a molecule covalently bound to Cys290, which restricts the movement of Tyr223 to the up conformation. Tyrosine 223's rotameric placement is also critical for active site conformation stabilization through the steric interference with Arg187. In the active form, the arginine 187's nitrogens are within 4Å of the tyrosine ring in the down conformation.

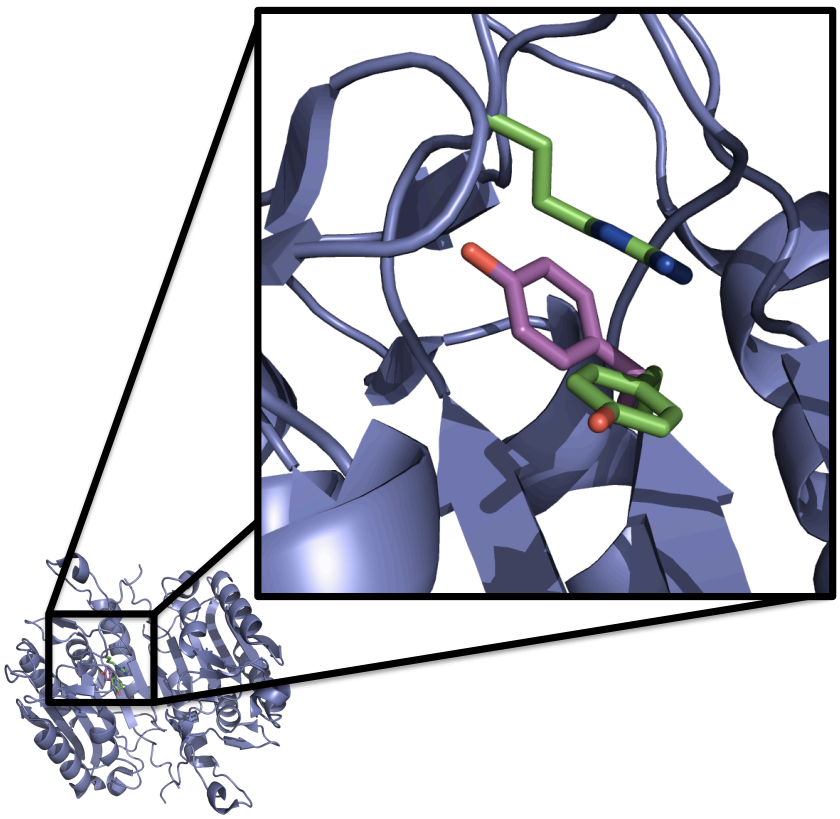


Figure 5: The caspase-7 allosteric switch

During allosteric inhibition with a small molecule bound within the cavity at the interface of the two small subunits (allosteric cavity), tryrosine 223's 'up' position (magenta) restricts arginine 187 from adopting the active conformation by forcing arginine 187 and L2 into a conformation not amiable to substrate binding. . In the active conformation (green), arginine 187 from L2 adopts a conformation that sterically disallows the 'up' conformation of tyrosine 223.

This arrangement suggests an interaction between the aromatic ring and the arginine. The arginine needs to move toward the cavity to transition from the inactive pro form, to the active form. This move is stabilized by the interaction, thus, the tyrosine's placement is critical.

In addition to R187, positioning of the L2' loop is essential for allosteric regulation. It should be therefore possible to emulate this inhibited conformation of L2' down over the allosteric cavity with a non-covalently bound small molecule, whose contact with the cavity will be sufficient to force Tyr223 into a position unfavorable for catalysis, or to

block the interaction of Arg187 with Tyr223 which will block the formation of the active site. Similarly, a variant of caspase-7 might be designed with restraints on loop movement creating an allosterically controllable protease.

Computational Protein Redesign

The goal of the Hardy lab is to develop additional vectors of control into proteins in novel ways using a rational design approach. By using data from x-ray crystal structures as a starting point, computational, rational or directed evolutionary approaches can be employed to systematically tune a protein for a desired response. Protein redesign whether by computational, rational or directed evolution methods is not a novel idea. Early efforts to use computers to augment structural information by way of homology modeling started in the early 1980s[43,44]. Efforts to improve drug-like properties of biological molecules have also benefited from computational design[45,46]. Additionally, the use of protein engineering to stabilize complexes for crystallography and functional studies is widely employed[47,48]. Computational protein design and redesign have become very powerful tool to study not only novel folds, but also to engineer novel functions. Efforts to computationally redesign proteins have met with mixed success[49].

Computational protein design has benefited immensely from recent advances in computing power, and thus seminal papers have been published on this topic. Early efforts of protein design[50-52] focused on a simplified problem that only targeted the hydrophobic cores. This pioneering work was necessarily less complex, but nonetheless established many algorithms[53,54] and prompted work on large combinatorial methods that followed[55]. This work emboldened protein chemists to use computational methods as the computational results were becoming more reliable. Two general approaches have been

tried in this field, *de novo* design, and using an existing protein as a scaffold for design. Mayo and colleagues have successfully used an automated search method to design a novel amino acid sequence that folds similar to a zinc-finger domain in the absence of zinc. Baker and colleagues, of Rosetta[56] fame, were able to design a globular protein of a fold that does not exist in nature *in silico*, and then solve the crystal structure to demonstrate the accuracy of their approach [57]. Not satisfied with strictly *de novo* design, Baker and colleagues have successfully reengineered endonuclease specificity[58], and built an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction[59] using similar methods. This substantive success in the field of protein redesign illustrates the effectiveness and more importantly, credibility of the method.

In this thesis, we examine the role of the loops of caspase-7, and their contributions to the formation of the substrate-binding pocket. Using these data, we harness the mobility of the L2' loop to engineer a redox activatable caspase-7. Finally, we explore the possibilities of allosteric cavity redesign to bind a small molecule using computational methods, employing data and knowledge gathered from the investigations of caspase-7.

CHAPTER II

LOOP L2' IS CRITICAL FOR ACTIVE SITE FORMATION

The majority of this chapter has been published: Witkowski, Witold. and Hardy, Jeanne A. "L2' loop is critical for caspase-7 active-site formation." Protein Science, 2009, 18, 1459-1468.

Abstract

The active sites of caspases are composed of four mobile loops. A loop (L2) from one half of the dimer interacts with a loop (L2′) from the other half of the dimer to bind substrate. In an inactive form, the two L2′ loops form a cross-dimer hydrogen-bond network over the dimer interface. Although the L2′ loop has been implicated as playing a central role in the formation of the active-site loop bundle, its precise role in catalysis has not been shown. A detailed understanding of the active and inactive conformations is essential to control the caspase function. We have interrogated the contributions of the residues in the L2′ loop to catalytic function and enzyme stability. In wild-type and all mutants, active-site binding results in substantial stabilization of the complex. One mutation, P214A, is significantly destabilized in the ligand-free conformation, but is as stable as wild type when bound to substrate, indicating that caspase-7 rests in different conformations in the absence and presence of substrate. Residues K212 and I213 in the L2′ loop are shown to be essential for substrate-binding and thus proper catalytic function of the caspase. In the crystal structure of the I213A mutant, the void created by side-chain deletion is compensated for by rearrangement of tyrosine 211 to fill the void, suggesting that the requirements of substrate-binding are sufficiently strong to induce the active

conformation. Thus, although the L2′ loop makes no direct contacts with substrate, it is essential for buttressing the substrate-binding groove and is central to native catalytic efficiency.

Introduction

Initiator caspases cleave executioner caspases in two locations to liberate an N-terminal propeptide and cleave the intersubunit linker. Executioner procaspases are dimeric so that cleavage of the intersubunit linker converts each half of the homodimer into one large and one small subunit. Mature caspases then consist of four chains, with two large subunits flanking two small subunits in the mature state. Cleavage of the intersubunit linker results in the formation of two new termini that function as loops involved in substrate binding. The C-terminal end of the large subunit becomes the L2 loop, whereas the N-terminal end of the small subunit becomes the L2′ loop. The catalytic cysteine is contained on the L2 loops, thus each half of the dimer contains one catalytic site[60]. The catalytic sites are surrounded by extended substrate-binding grooves. Each substrate-binding groove is composed of three loops (L2, L3, and L4) from one monomer. A fourth loop (L2′) from the opposite monomer interacts with L2 to stabilize the substrate-binding loop bundle in the active state (Figure 6 A,B).

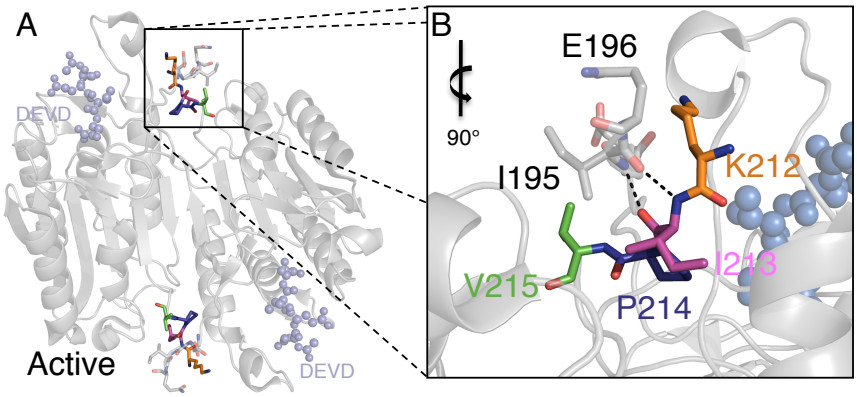


Figure 6: Active caspase-7

(A) The L2' loop (colored sticks) is in the ''up'' position only when substrate binds to the active conformation of caspase-7 (PDB ID 1F1J). Residues selected for mutation (K212, orange; I213, magenta; P214, blue; V215, green) make interactions with the L2 loop from the opposite half of the dimer. The two substrate-binding grooves are bound by inhibitor DEVD (light blue ball-and-sticks), which is a substrate- mimicking compound. Loops L3 and L4 contribute to the substrate-binding loop bundle with loops L2 and L2'. (B) Interactions between L20 (colored sticks) and L2 (gray sticks) stabilize the active form of caspase-7.

Large conformational changes in L2 and L2′ are associated with the transformation between the caspase-7 conformational states, all of which have been structurally characterized. When the inactive procaspase-7 zymogen[39,41] is cleaved to generate mature caspase-7,[39] it remains in a conformation that has been observed crystallographically that is between active and allosterically inhibited ("semi-active") until substrate binds, inducing the active caspase-7 conformation[37,38,61]. Caspase-7 can also be allosterically inhibited[42] by binding of effectors to a cavity at the dimer interface. In the inactive procaspase-7 zymogen L2 and L2′ are covalently connected yet they are observed to be pointing away from one another, separated by a stretch of disordered residues. This conformation is very similar to that observed in the allosterically inhibited state, where the L2′ loops interact above the allosteric site (Figure 7 C,D).

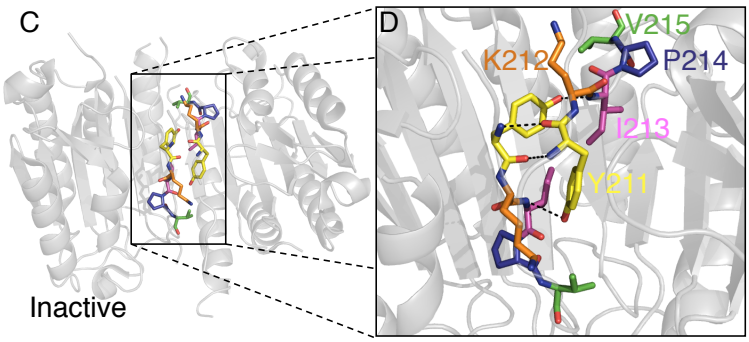


Figure 7: Allosterically inhibited caspase-7

(C) The L2' loop (colored sticks) is in the ''down'' position both in the zymogen and allosterically inhibited (allosteric model shown, PDB ID 1SHJ) conformations. The magnitude of the L20 loop rearrangement is clearly visible compared to (A). Mutated residue Y211 is represented as yellow while resides 212–215 are colored is as in (A). (D) The hydrogen-bonding network (dashed lines) at the dimer interface comprises interactions between L2' from one monomer and L2' from the opposite side of the dimer (dashed lines).

In the absence of substrate, the mature caspase-7 is in a partially active conformation (Figure 8 E,F). L2 is in the conformation that allows substrate binding, but L2′ is in the down position that is more similar to the zymogen and allosterically inhibited conformations. Thus, the L2′ position correlates with caspase-7 activity.

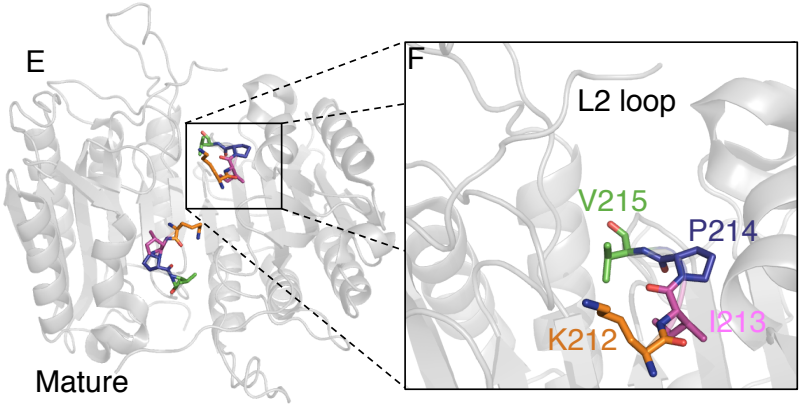


Figure 8: Mature caspase-7

(E, F) The half-active conformation observed in the mature (cleaved but substrate-free) form of caspase-7 (PDB ID 1K86) is colored as (A). Although the L2 loop is in a similar conformation to the active-site-bound structure, the L2' loops point down toward the allosteric cavity in a conformation most similar to the allosteric/zymogen conformation. For additional clarity not possible in paper form, see Supporting Information Movie S1.

This canonical caspase-7 activation model suggests L2 and L2′ can only interact properly after cleavage by initiator caspases to generate the free loops, meaning that both of the caspase monomers must be cleaved to achieve a fully active caspase. Thus, both the L2 and L2′ loops are central in caspase activation. Recently, however, hybrid dimers of one uncleavable and one cleavable caspase monomer suggest that cleavage of both monomers in not necessary for formation of one catalytically active site[62,63]. This implies that the uncleaved intersubunit linker can also adopt the conformation of either L2 or L2′ but not both. The L2 loop has been implicated in an allosteric network in caspase-1 in which the two catalytic sites are coupled through an extensive hydrogen-bond network.[62] It remains to be seen if this allosteric coupling is present in all of the caspases and what role the L2′ might play in this mechanism.

The fact that the L2′ loop is not in the correct orientation in the mature and zymogen conformations has been suggested to be the basis for the lack of activity in the zymogen[39]. Some mutagenesis inquiries have probed the role of L2 in caspase activity, and report that mutations in this region are deleterious.[64,63,32] On the basis of structural observations, we anticipate that the L2′ loop likewise plays an important role in caspase function. To our knowledge, we report the first systematic study of the L2′ loop. Given the difficulties that have been encountered in developing active-site-directed caspase therapeutics, this region is particularly interesting because it is distal from the catalytic site, but appears to be influential for both substrate binding and catalysis.

Analysis of the existing crystallographic structures of caspase-7 indicates that residues in the L2′ loop buttress the L2 loop into the proper configuration of the active-site loop bundle. We investigated the role of the L2′ loop by serial alanine substitution of residues

211–215, which form the core of the buttress. The resulting mutants were tested to determine the severity of these mutations on kinetics and apparent thermal stability. To assess the structural changes in the most severely affected mutant, we solved the crystal structure and compared the results with the existing caspase-7 structures.

Results

Role of the L2′ loop in catalytic efficiency in caspase-7

The placement of L2′ in the active conformation at the base of the active-site loop bundle suggests that it may be essential for substrate binding. Alanine mutants were designed to probe the regions of greatest interaction between L2 and L2′ (211–215) in previous crystal structures of caspase-7 in the active conformation (Figure 6 A).[37,38,61] In caspase-3, the residues in the region of the L2′ loop are identical to caspase-7 (Figure 9) with the exception of position 211 (caspase-3 position 185). Additionally, in the crystal structures of active caspase-3, the L2′ loop is in an identical conformation[29,37,65,31,33,66] suggesting that the buttressing role is conserved within the executioner caspases.

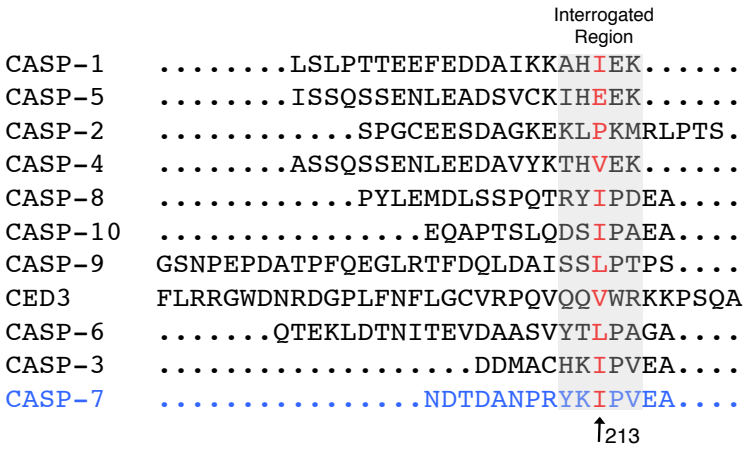


Figure 9: Sequence alignment of the L2' loop in related caspases

Caspase-7 residue 213 (Ile) is conserved in size or type in not only apoptotic executioner caspases (caspase-6, -3, and -7) but also apoptotic initiators (-8, -9, and -10), *C. elegans* CED3, the first discovered caspase, the inflammatory caspases (-1 and -4) and caspase-14, the epidermal-specific caspase. This residue is not conserved in the initiator caspase-2 or the inflammatory caspase-5.

Given the buttressing role of the L2′ loop, alanine substitution should be a useful probe for the role of each residue in the formation of the active-site loop bundle as mutations of critical residues would have a strong impact on K_m.

Kinetics of wild-type and mutant caspase-7 variants were tested as a function of substrate concentration to evaluate K_m and k_{cat} (Table 1).

	K_m (μM)	k_{cat} (s^{-1})	10^6 x [k_{cat}/ K_m] ($M^{-1}s^{-1}$)	Relative Activity (relationship factor of mutant k_{cat}/K_m versus WT)
WT	23 ± 2	$0.35 \pm 3\times10^{-3}$	0.015	1.0
Y211A	18 ± 10	$0.39 \pm 3\times10^{-3}$	0.021	1.4
K212A	210 ± 20	$0.93 \pm 4\times10^{-2}$	0.004	0.26
I213A	480 ± 50	$0.07 \pm 5\times10^{-3}$	1.5×10^{-4}	0.01
P214A	8 ± 1	$0.10 \pm 4\times10^{-3}$	0.012	0.8
Y215A	26 ± 8	$0.70 \pm 2\times10^{-2}$	0.026	1.7

Table 1: Kinetic parameters for caspase-7 alanine-substitution mutants

Michaelis-Menten kinetic parameters measured on L2' alanine-substitution mutants demonstrate that positions K212 and I213 are most affected by the alanine substitution.

As anticipated, mutation of two of the residues, K212 and I213, had a significant effect on substrate binding as assessed by K_m, while the K_m for K212A was ~10-fold weaker than wild-type caspase-7. The K_m for I213A at ~20-fold weaker than wild type is the most significant effect observed. I213A is the only mutation that has a significant impact, an ~5-fold decrease, in k_{cat}. Mutation at Y211, P214, and Y215 had minimal impact on the catalytic efficiency. Catalytic efficiency (k_{cat}/K_m) of only K212A and I213A was significantly different from wild type, at 4- and 100-fold decreases, respectively. These results identify K212 and I213 as the most critical residues in the L2′ buttress.

Alignment of the L2′ region of the caspases (Figure 9) indicates hydrophobic and size conservation of the 213 position, but not a strict conservation of amino acid identity. Caspase-1, -3, -7, and -10 use isoleucine at this position, whereas other caspases substitute only valine and leucine, which are nearly isosteric with isoleucine. Position 212, which exhibited the second most significant K_m effect when mutated to alanine, is not conserved across the caspase family. Position 212 is nearly always hydrophilic;

however, sizes vary from small (serine) to long (lysine) or bulky (histidine and tyrosine). In the active conformation of wild-type caspase-7, the exposed aliphatic portion of the lysine side chain is packed against the aliphatic portion of E196 (Figure 6 B). Interestingly, these residues are not strictly conserved in the third executioner caspase, caspase-6. Conservation of this region is weak compared with the region flanking the catalytic cysteine C186 (residues 178–188).

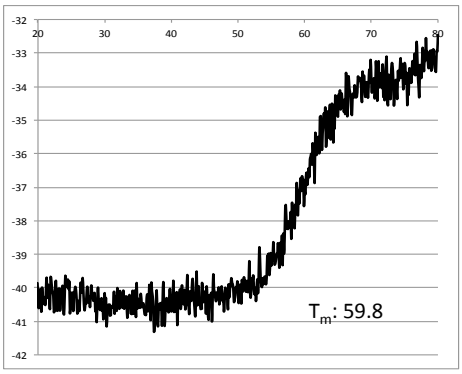
L2′ mutations affect protein stability

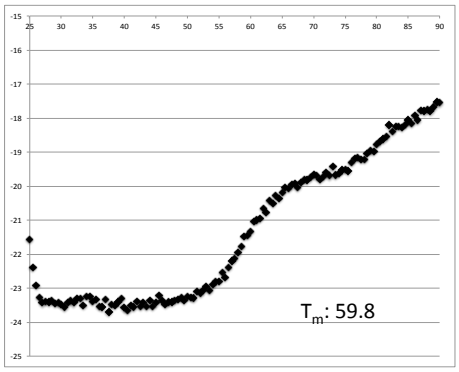
Alanine mutations have been reported to generate cavities in protein cores[67]. Because alanine mutants have the possibility of introducing a destabilizing cavity, we probed the stability of the protein in the mature (substrate unbound) form using thermal denaturation monitored by circular dichroism (Figure 11). Inspection of caspase-7 structures suggests that the only important interactions that form in the substrate-bound conformation but not in the mature or zymogen form are between L2 from one monomer and L2′ from the opposite monomer. Thus, interrogation of the stability of wild-type and mutant caspases in the active conformation [with substrate-like inhibitor, the caspase-binding peptide aspartate–glutamate–valine–aspartate (DEVD) bound] was also relevant.

The serial alanine mutations appear to have no discernable effect on the overall fold of the protein in either the inhibitor-bound versus unbound states, as the measured spectra at 25°C were unaffected by mutation. The melting temperature (T_m) of all mutants in the unbound form was similar to wild type at ~60°C, except for the P214A mutation which had a T_m of 46°C (Figure 11).

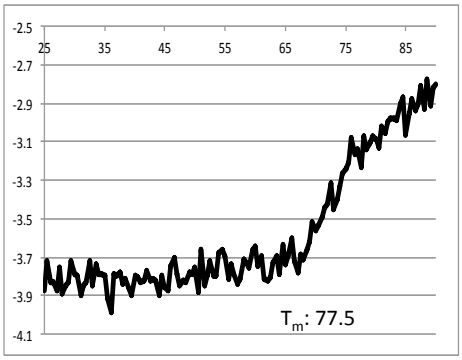
C7 WT

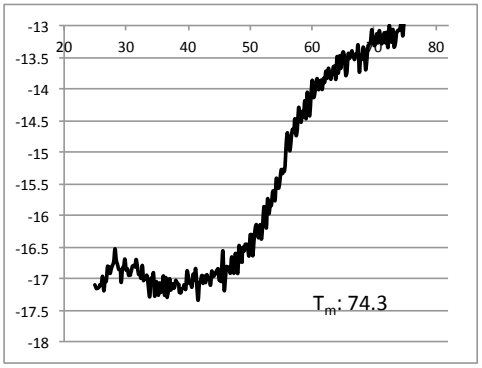
APO





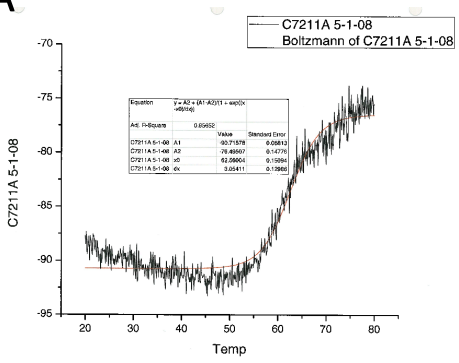
+ DEVD-CHO

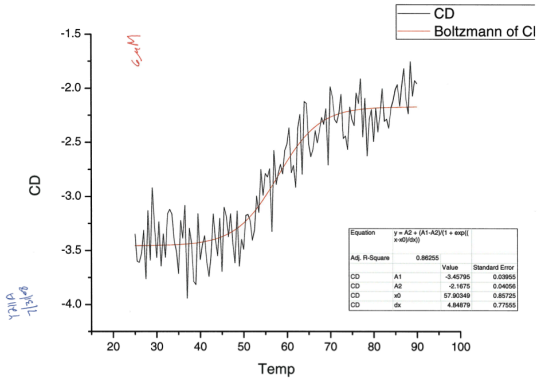




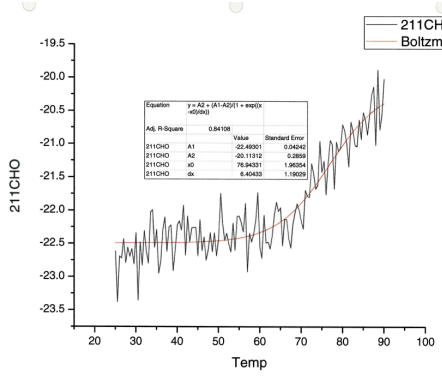
Y211A

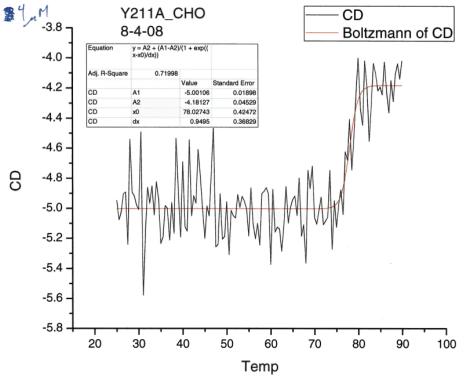
APO





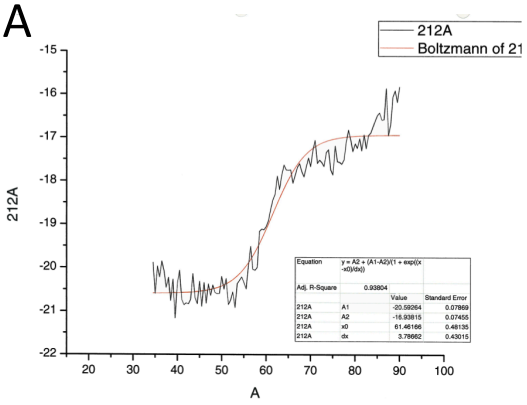
+ DEVD-CHO

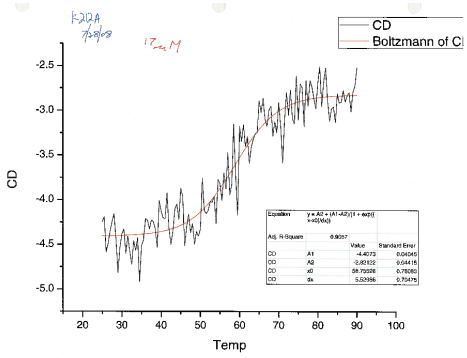




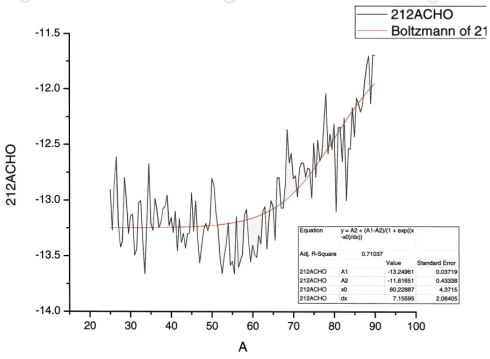
K212A

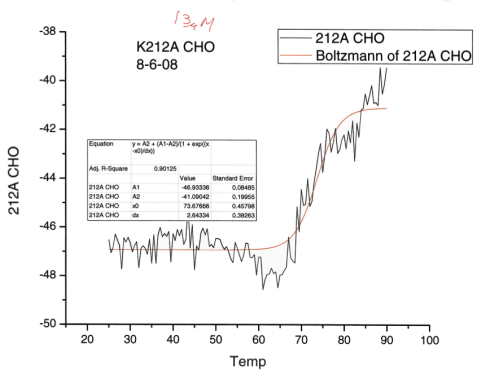
APO





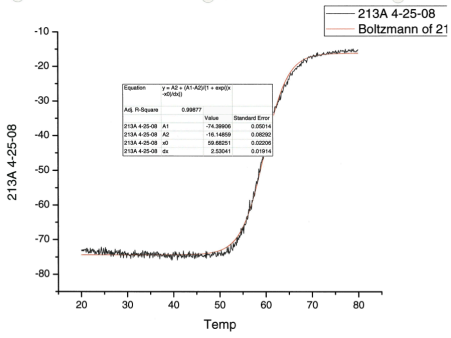
+ DEVD-CHO

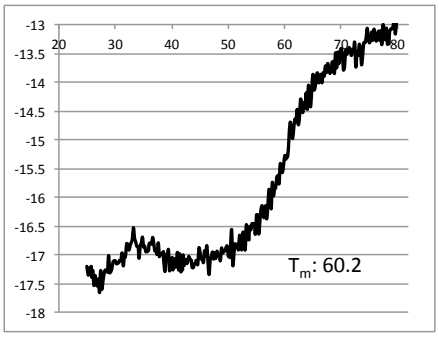




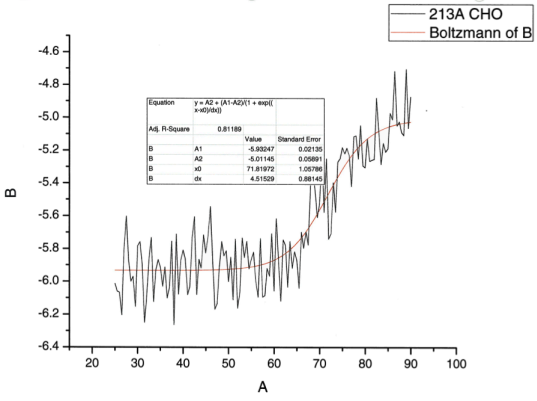
I213A

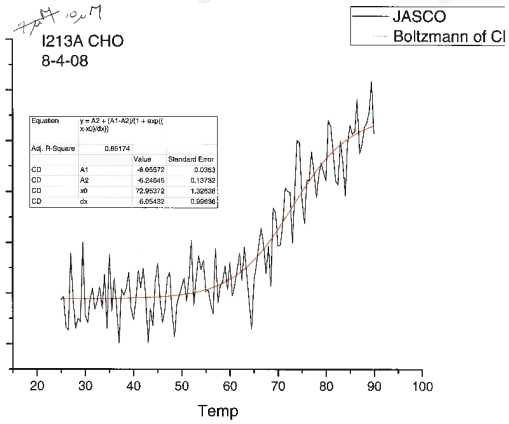
APO





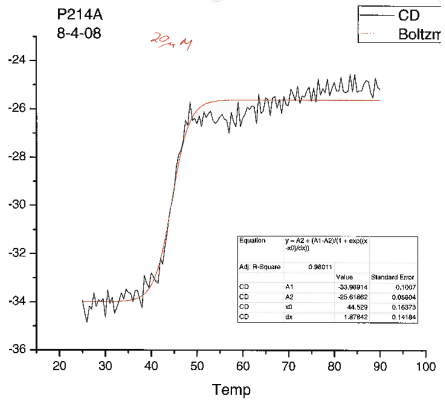
+ DEVD-CHO

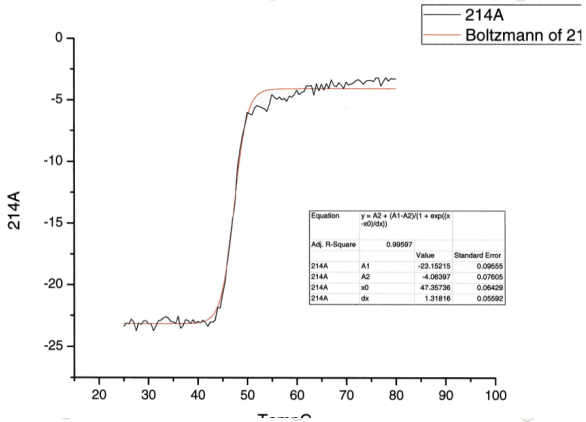




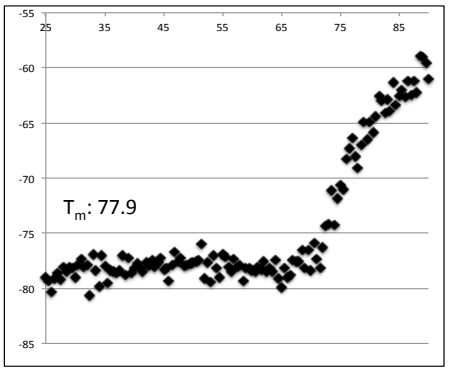
P214A

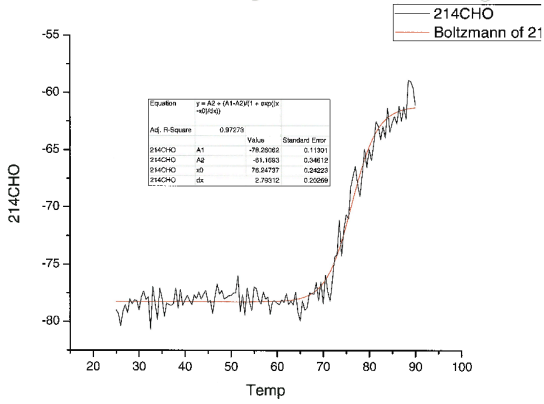
APO





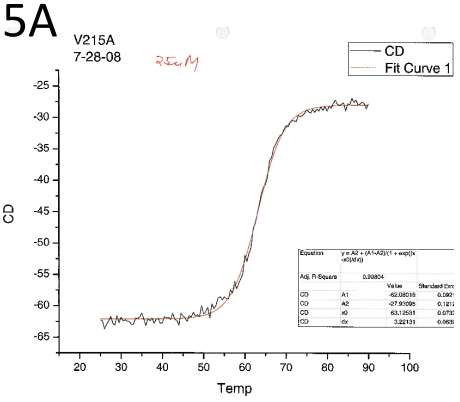
+ DEVD-CHO

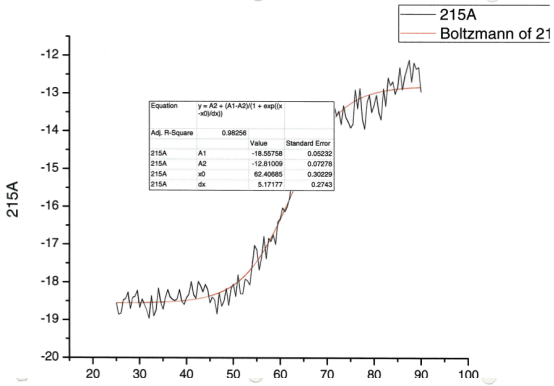




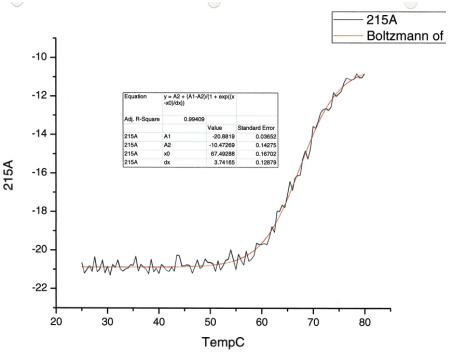
V215A

APO





+ DEVD-CHO



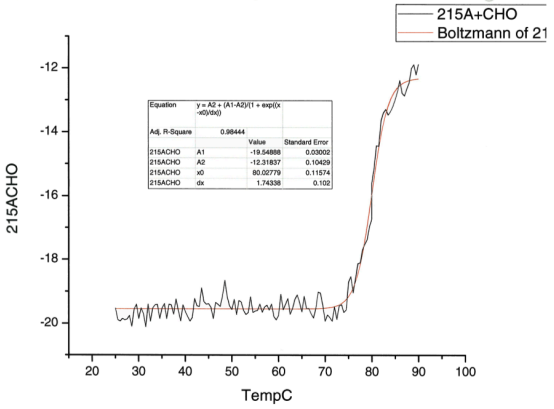


Figure 10: T_m Curve Data for WT and alanine scanning mutants

Raw circular dichroism data monitored at 222 nm during a thermal denaturation of proteins. T_m calculated by fitting a Boltzmann sigmoidal curve to data collected. Data for V215A, K212A+DEVD-CHO and WT+DEVD-CHO were collected in triplicate, not shown. All other data were duplicate measurements made independently on different preparations on separate days.

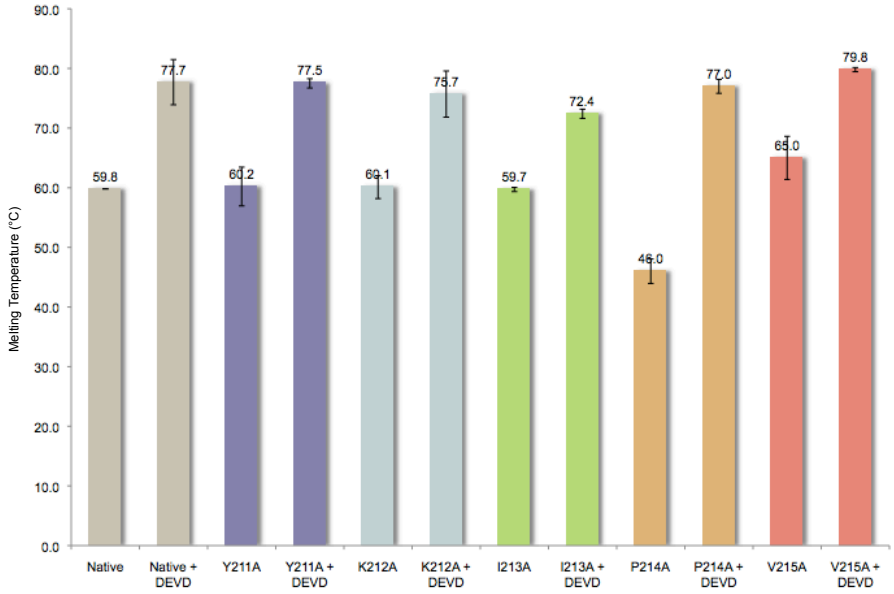


Figure 11: Melting temperatures and cavities generated in caspase-7 variants.

Melting temperatures (T_m) were measured by thermal denaturation monitoring the change in CD signal at 222 nm in the presence (white bars) and absence (black bars) of the active-site inhibitor peptide DEVD-CHO. Measurements were made at least in duplicate with independently prepared samples on separate days. Error bars represent standard deviation between replicates measured on separate protein samples on different days.

We had hoped to also measure thermodynamic stabilities of the caspase-7 mutants; unfortunately, unfolding of wild type and all of the mutant caspase-7 variants was irreversible and precluded accurate assessment of $\Delta G_{unfolding}$. We calculated the predicted volume changes each mutation could contribute to the overall structure (Figure 12) of caspase-7 in the active conformation assuming there were no other structural changes in the protein, either backbone or side chain position. From these figures, we were able to assess the potential volume of each deletion and what role it might play in the stability of

each protein. The Matthews group has systematically assessed the effect of cavity-forming mutations in the core of T4 lysozyme. In one study of 44 large amino acid to alanine substitutions, cavities between 17 and 123 $Å^3$ were observed crystallographically.[67] These cavities correlated with decreases in stability ($\Delta\Delta G$) of 0.9–5.0 kcal/mol. Several cavities were observed in the range of 25 $Å^3$ that had $\Delta\Delta G$ ranging from −3.2 to −2.7 kcal/mol. In T4 lysozyme, a 1.9 kcal/mol change in $\Delta\Delta G$ corresponds to approximately a 5° decrease in T_m.[68] Thus, the 14° decrease in T_m of unbound P214A could reflect a $\Delta\Delta G$ of greater than 4 kcal/mol.

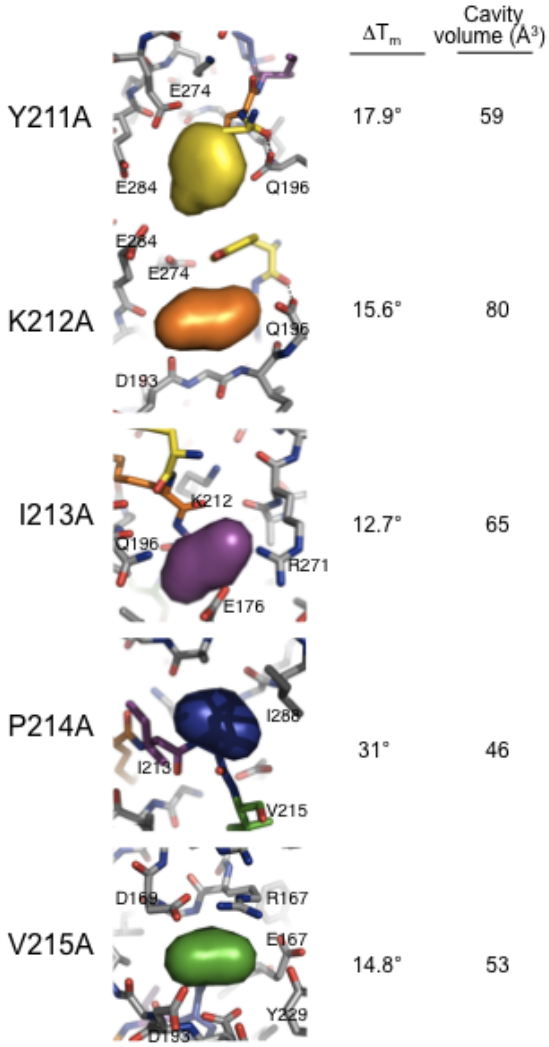


Figure 12: Predicted volumes of cavities in models of mutant proteins

Volumes of cavities in models of mutant proteins generated computationally in PyMol were assessed with Pocketfinder. The deleted volumes are shown as colored pillows to illustrate the chemical environment and residues forming the generated cavity. Change in the T_m of caspase- 7 dimers in the presence of active-site-binding peptides (DEVD-CHO) relative to ligand free is also listed.

Interestingly, the T_m difference between mature and active-site-bound forms was pronounced, with an average 18.7°C increase in T_m of the active-site-bound form over the unbound form. The single observed structural difference between these two forms is that the inactive, unbound form is able to sample both the up conformation of L2′ and the inactive down conformation of L2′. When substrate is bound, L2′ is locked in the active

up conformation. Thus, when the protein adopts the up, bound form, it appears to be significantly stabilized by the interactions between L2′ and L2 from opposite monomers and by interactions with substrate.

I213A has a smaller than average ΔT_m of 7°C between the forms (Figure 11), indicating that an energetic penalty must be paid to adopt this conformation because the active form is less stable to thermal denaturation. The dramatically destabilized P214A mutant, on the other hand, showed a ΔT_m of 31°C between bound and unbound forms. Although the identity of 214 seems to be more influential in the inactive form, in the active form, the stability is comparable to the wild-type enzyme. This can be explained by structural examination of the active form where P214 seems to have a more limited role in buttressing than I213 (Figure 6, Figure 7, Figure 8). In the active conformation, L2–L2′ physical interactions tend to dominate over any geometric constraints imposed by P214. The substitution of P214A has a negligible effect on active (bound) enzyme stability; however, in the inactive form, the hydrogen-bonding network present between L2′ and L2′ from the opposite monomer would likely be disrupted as the geometric constraints imposed by the proline at position 214 are removed by the alanine substitution. This substitution, however, does not inhibit dimerization of the protein, as measured by size exclusion chromatography (Figure 13). Both WT and P214A proteins ran at an apparently smaller size than expected for the dimer of caspase-7 (60 kD). We noted a decrease in apparent size of P214A relative to WT, however we postulate that the more flexible nature of the loops decreases the proteins apparent interaction with the gel, causing the smaller than expected (60 kD) size.

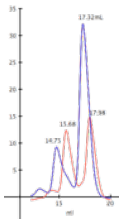
Protein		RT (mL)	SEC Observed MW (KDa)	Expected MW (KDa)
Standards				
P214A		15.68	45	60
P214A + DEVD		15.26	55	61
WT		15.28	54	60
WT + DEVD		15.33	53	61

Figure 13: Size exclusion chromatography

Size exclusion chromatography of caspase-7 WT and P214A with and without inhibitor bound. In the presence of inhibitor, P214A's calculated size is similar to WT with inhibitor. The difference in size of apo P214A versus WT is not appreciable.

Analysis of caspase-7 by analytical ultracentrifugation

We postulated that the decreased stability of caspase-7 P214A due to the more flexible L2' loop would have an appreciable effect the dimerization constant. Clark and colleagues have reported an estimation of the dimerization constant for caspase-3 to be below 50 nM[69], so we imagined that caspase-7 P214A would be sufficiently destabilized in the apo state to monitor this by analytical ultracentrifugation. P214A in the absence of

covalent catalytic inhibitor was studied by analytical ultracentrifugation in equilibrium mode to determine dimerization constant. The estimated molecular weight from AUC was determined to be 53 kD and a single species. Fitting monomer-dimer equilibrium was unsuccessful as at experimental conditions of 10-20 μM, no monomer was detected. These data suggest that the K_d of dimerization is significantly below 10 μM and does not appear to be changed from WT caspase-7. Thus, further experiments to determine experimental K_d of WT by AUC was not attempted due to equipment restraints. In hindsight, the ability to collect accurate kinetics on protein at 100 nM should have suggested that the P214A's dimerization constant was below 100 nM.

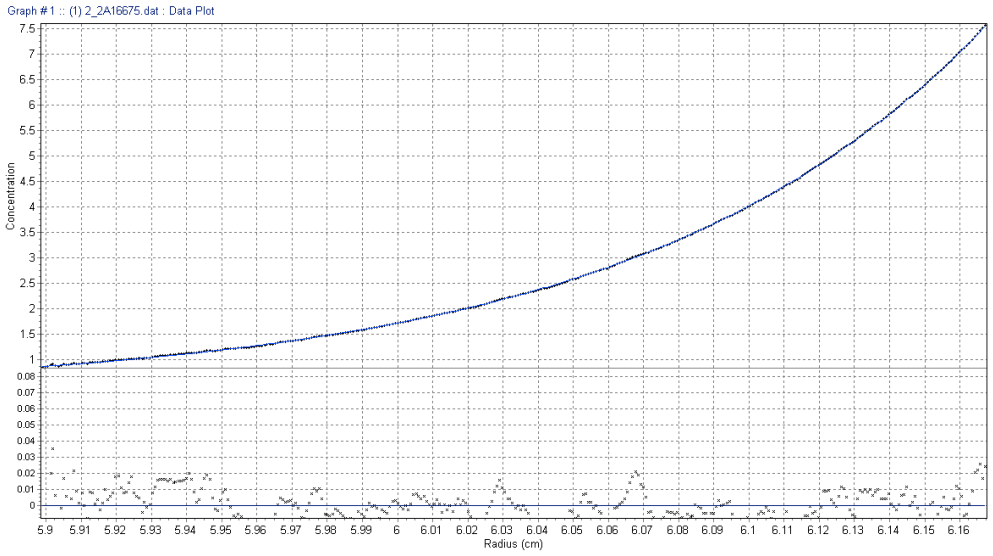


Figure 14: Equilibrium AUC analysis of P214A apo

Single species fitting of equilibrium data based on two concentrations and three rotor speeds. Fitted line (blue dots) overlays data (black dots) very well, with statistical residuals (bottom) maintaining sufficient randomness.

Additional experiments undertaken to lower experimental concentrations and thereby increase monomer population in solution were unsuccessful, because the limit of detection of the instrument was reached. At concentrations below 5 μM protein, data

collected were too noisy to process, making measurements in the concentration regime of caspase dimerization impossible[69]. Thus we can only conclude that the dissociation constant is well below 10 μM. Furthermore, because caspase-7 and the P214A mutant show activity at 100 nM protein concentrations, and dimerization is known to be essential for activity, we can further conclude the dissociation constant is less than 100 nM in the presence of substrate.

```
MW = 52558 using the current values of vbar = 0.7300 and density = 1.00000
Sigma at Max. Speed of 21000 RPM = 2.816 cm-2

Number of Iterations: 5
Number of Points: 2084
Degrees of Freedom: 2071
RMS Deviation: 0.01118
Number of Parameters Held Constant: 0

Variance Ratio at the 95.46% confidence level:
F(NDF,NDF,1-P) = 1.0772
```

Figure 15: MW calculation for P214A apo AUC data

Fitting parameters and results from single species fitting of equilibrium data based on two concentrations and three rotor speeds. A molecular weight of 53kD was fit with 2084 points, with an RMSD from fit of 0.011.

Thus, we can conclude that substrate binding and the concomitant stabilization of interactions observed between the L2 and L2′ loops are the most critical factors in significant stabilization of the active-bound conformation. Indeed, substrate binding appears to be so influential in locking L2′ securely into the up conformation that it can overcome significant thermodynamic destabilization in the unbound state. Because P214A had negligible effect on both the stability of the substrate bound form and the kinetics of the enzyme, we focused our investigation on the structural differences caused by the I213A mutation to further understand the structural role of the L2′ loop in substrate binding and catalysis.

P214A Mutation does not affect DICA binding

In the crystal structure of caspase-7 bound to allosteric inhibitors (1SHJ and 1SHL), loop L2' is positioned over the allosteric site. In those structures, residues I213 and Y211 form a hydrogen bonding network that potentially stabilizes the resting state or 'half active' conformation of caspase-7. Stability measurements (Figure 11) suggest that in this hydrogen bonding network is disrupted by the mutation of P214 to alanine, likely due to the additional degrees of freedom afforded by the alanine. Proline imposes geometric restraints not found in alanine, and the subsequent destabilization of the loops could account for the 13.8° difference in T_m. To investigate whether this difference and loop flexibility would affect DICA binding and removal, caspase-7 WT and P214A were both incubated with DICA. Both proteins responded similarly to DICA inhibition, and were found to be fully inhibited within 1 hour. Reductive removal of DICA from the allosteric cavity using DTT at various concentrations (100 μM and below in a two fold dilution series) indicated no statistically significant difference between the wild type enzyme and the P214A variant. (Figure 16) This suggests the loop L2' makes additional interactions with DICA that stabilize L2' in the down conformation, occluding the accessibility of reductant to C290 in both WT and P214A. Interestingly, it was found that high concentrations of TCEP have an inhibitory effect on both WT and P214A caspase-7 at concentrations above 50 μM. This same effect was not observed with the use of DTT.

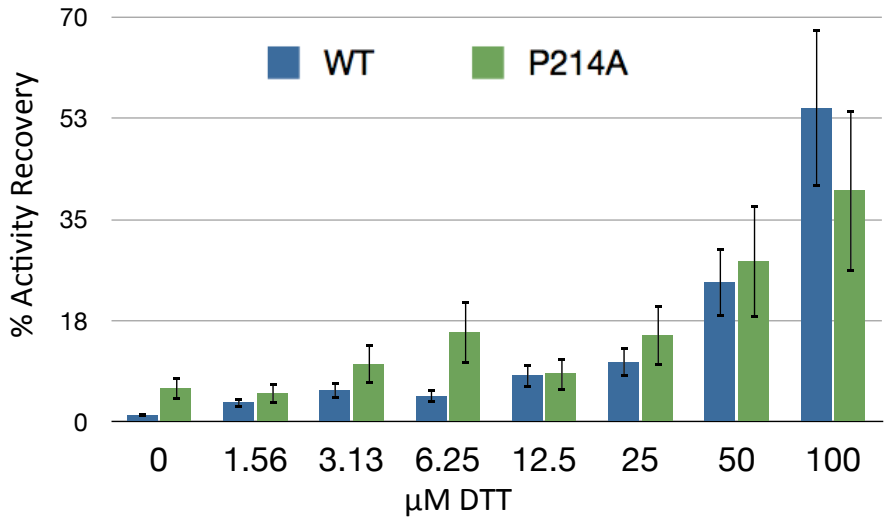


Figure 16: Percent recovery of activity as a function of DTT concentration

Caspase-7 WT and P214A do not show statistically significant differences when inhibited with allosteric disulfide linked DICA, followed by treatment with DTT reductant. Activity returns as a function of reductant concentrations to both WT and P214A.

Crystallographic studies of I213A in the presence of active-site inhibitor

Crystals of caspase-7 I213A with DEVD covalently bound in the active site to lock the protein in to the active conformation were grown in a variation on known conditions,[61] and the structure was solved at 2.6 Å (Table 2). To avoid model bias during structure determination, we omitted all of the active-site loops from the molecular replacement search model and thus calculated an omit map that is unbiased in the loop region as our initial map (Figure 17 A,B). Compared to an active caspase-7 structure that crystallized in the same form (1F1J[37]), the RMSD of all atoms was found to be 0.191 Å, further supporting the observation that the I213A mutation had no effect on the overall fold of the protein. The substrate-binding grooves and active-site inhibitor conformations in our structure were virtually unchanged from wild-type caspase-7.

Diffraction Data	
Wavelength (Å)	1.54
Resolution Range (Å)	50-2.61 (2.70-2.61)
Measured reflections	26499
Unique reflections	10599
Completeness (%)	97.8 (92.0)
Redundancy	2.5 (2.5)
$\langle I/\sigma_I \rangle$	22.5 (1.99)
Space Group	$P3_221$
a = b (Å)	89.4
c (Å)	186.1
α = β (°)	90
γ (°)	120
R_{sym}	8.3 (69.4)
Refinement Statistics	
Atoms	3919
Water molecules	119
R_{work} (%)	19.9
R_{free} (%)	25.1
RMSD bond length (Å)	0.019
RMSD bond angle (°)	1.876
Average B-factor (Å)	51.82

Table 2: Statistics for the x-ray crystal structure of caspase-7 I213A

Data in parenthesis are for the highest resolution bin

The greatest difference between the wild-type and the I213A structure was in the Y211 region (Figure 17 A–C). Y211 is unresolved in most caspase-7 structures of the active form, and only recently has been observed in structures (PDB ID 2QLF and 2QLJ) of active caspase-7 bound to peptide inhibitors with aldehyde warheads[61].

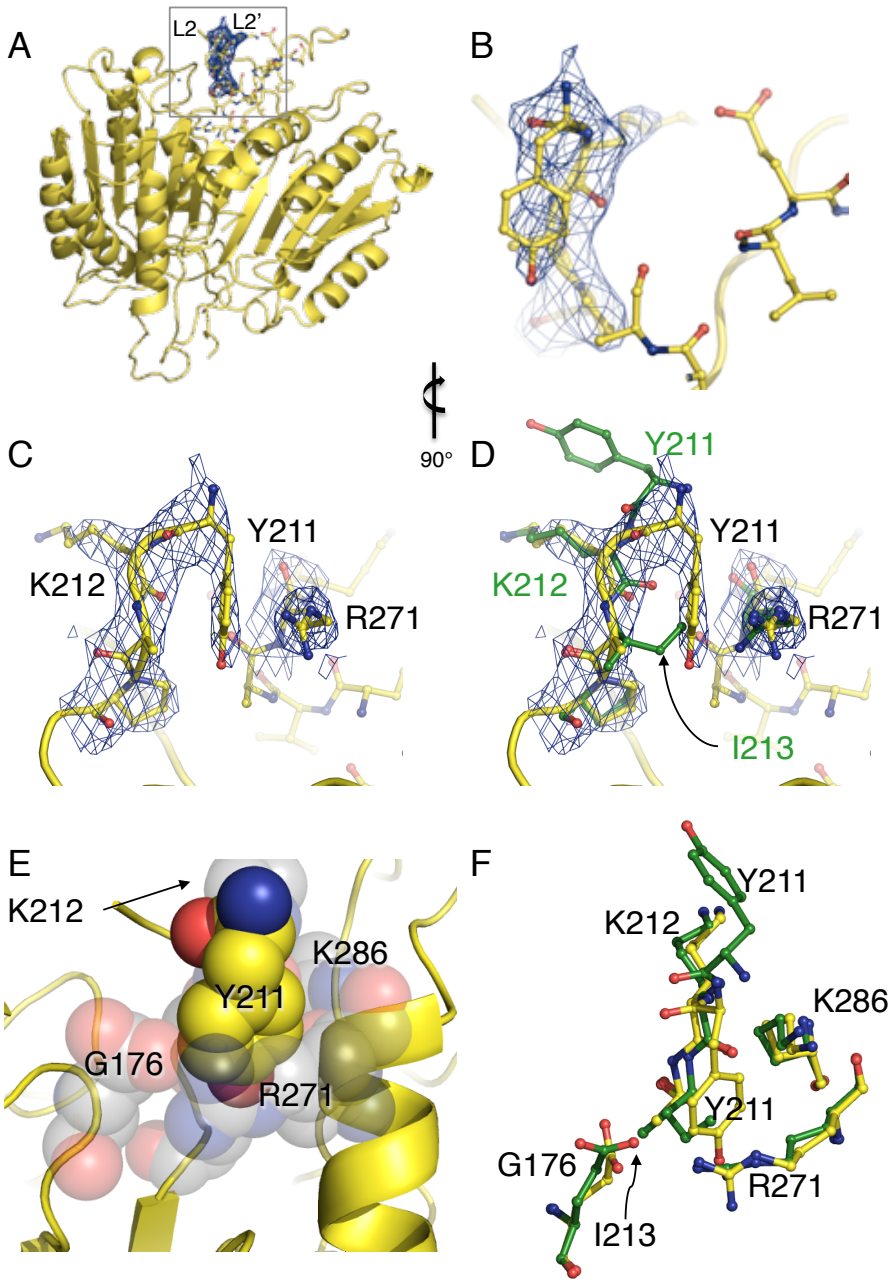


Figure 17: Crystallographic evidence of compensatory mechanism that facilitates active-site-binding in caspase-7 I213A.

(A) Caspase-7 I213A structure with buttress region highlighted in inset. (B) $2F_o$-F_c density (blue mesh) from the initial omit map (residues 211–215 omitted from the phase calculation) and final refined model of I213A (yellow sticks) shown in same orientation as (A). (C) is as (B) but rotated 90° for clarity. The unbiased omit map clearly indicates the down position for Y211. (D) Wild-type caspase-7 (green sticks) is superimposed with the final I213A model, demonstrating that the conformation of Y211 fills a structural void when the I213 side-chain is deleted. (E) The environment (gray spheres) into which Y211 (yellow spheres) inserts is hydrophilic in nature to accommodate the hydroxyl moiety. (F) Residues forming the cavity into which Y211 insert are in a nearly identical conformation in wild-type caspase-7 (green sticks) and I213A (yellow sticks).

In both these structures, Y211 is solvent exposed (Figure 17 D,F). In our structure (yellow), Y211 exhibits a radically different conformation, burying itself into a newly formed surface cavity. This burial appears to be a compensatory mechanism for the I213A cavity-forming mutation (Figure 17 D,E). This position of the tyrosine fills a void in the region, where I213 is found in the wild-type structure. In the wild-type conformation, Y211 burial is sterically occluded by I213 (Figure 17 E).

This structural observation helps to explain the diminished k_{cat}, in the context of residual catalytic activity in the I213A mutant. Given that the apparent thermal stability of I213A in the active-site-bound form is lower than any of the other mutants, and the lack of any other structural changes in this mutant crystal structure, we can conclude that the cavity formed by deletion of the I213 side-chain is so unfavorable as to prevent proper formation of the loop bundle without some compensatory reorganization. The compensatory conformation we observe allows the formation of a catalytically competent enzyme but not without an energetic penalty.

Discussion

Based on the work presented here, two main points emerge. First, binding of substrate stabilizes caspase-7 via a structurally validated mechanism. Second, the core of the L2′ loop is residue I213 and this buttressing region is an essential component of active caspase-7. If the core of the L2′ loop is perturbed, then the L2′ loop does not properly hold L2 in place, the substrate-binding loop bundle is therefore destabilized and K_m of the resulting mutants are increased.

Caspase-7 is a constitutive dimer and is not believed to exist in the monomeric state to a significant degree[70]. In contrast, the initiator caspase-9 is predominantly an inactive monomer before activation by the apoptosome. Binding of active-site inhibitors drives caspase-9 to the dimeric state[70]. Thus, it is plausible that other caspases would likewise be stabilized by binding of active-site inhibitors. To our knowledge, no other group has shown and quantified the dramatic stabilization of the active-site bound form of caspase-7 compared with the unliganded state. Prior computational analysis of the stability of the active versus unbound conformation mirrors our experimental results, indicating that the L2′ up or active conformation of caspase-7 was significantly more stable than the L2′ down conformation[71].

The roles of individual residues in the L2′ loop obviously change depending on the state (active-site-bound or unbound) of the protein. The structures of the mature unbound form of caspase-7 (Figure 8 E,F) when compared with the active (Figure 6 A,B) and allosterically inhibited (Figure 7 C,D) forms suggest an ensemble of states that are in dynamic equilibrium with one another. Two of our mutants, I213A and P214A, support the idea that the L2′ loop is critical for stabilizing the active-site-bound conformation.

I213A caspase-7 has the same apparent thermal stability in the unbound form as does wild type. This indicates that the equilibrium constant for unfolding of the unbound form is likewise unchanged from wild type. On the other hand, the T_m of the active-site-bound form is slightly lowered, possibly suggesting that an energetic penalty must be paid to achieve this conformation. The crystal structure of I213A indicates that the decrease in stability is due to the cost of hydrophilic burial when Y211 intercalates into the I213A cavity to properly buttress the loop bundle (video of interaction can be accessed from

Protein Science Supplementary resource website, at: http://onlinelibrary.wiley.com/store/10.1002/pro.151/asset/supinfo/PRO_15_sm_suppmov2.mov?v=1&s=479f68a02900b7636d650589416d8e9a5bcb35ea). We anticipate that in the unbound form of I213A, L2′ is free to sample the down conformation, whereas in our structure, binding at the active site locks L2′ into the active, up, conformation. If caspase-7 natively sampled the active, up, conformation to a significant degree when the active site was empty, this up conformation would lower the stability of the unbound form of the enzyme. Because the I213A unbound apparent thermal stability is unchanged, it seems that the up conformation of the L2′ loop does not contribute significantly to the conformational equilibrium ensemble of the unbound form. Thus, proper ordering of the L2′ loop as a buttress for L2 is so essential for active-site-binding that the enzyme will pay an energetic penalty to maintain the active state.

The P214A mutant displays the opposite effect on stability. P214A is destabilized in the unbound state by a significant margin (14°C relative to wild type). Interestingly, when it is locked into the active conformation by active-site binding, the complex is as stable as the wild-type enzyme. This indicates the absence of energetic penalties like those observed for I213A and further suggests that the conformation of this mutant when the active site is bound is identical to wild type. The interactions of L2 with L2′ are the only additional interactions within the caspase protein itself that are observed on going from mature unbound caspase-7 to active-site-bound caspase-7. The fact that a dramatic destabilization of the unbound form of P214A can be overcome by active-site binding indicates that these L2/L2’ interactions are responsible for the jump in apparent thermal stability observed in all versions of caspase-7. Structural analysis suggests that the L2′-

loop up conformation should be more stable as it has better burial of exposed hydrophobic residues. Likely, an entropic penalty for ordering of mobile loops prevents this conformation in the absence of active-site binding. Moreover, the conformational equilibrium in the bound state has essentially no component of the unbound state, which would lead to destabilization of both the unliganded and liganded states.

One of the looming controversies in understanding caspase structure and activity is the structural state of the mature (cleaved) protein in the absence of active-site ligand. There is but one existing crystal structure of unliganded, mature caspase-7[39]. In this structure, the active-site loops do not adopt the substrate-bound conformation (L2 and L2′ up), but are in a half-active conformation with L2 up but L2′ down. In contrast, mature caspase-3 has been reported to exist in the liganded conformation (L2 and L2′ up) with an empty active site, crystallized in the presence of what the authors term a noncatalytic site inhibitor, which is not observed in the crystal structure.[31] The fact that a mutation in L2′ exists, namely P214A, that dramatically affects the stability of the unliganded state but not the liganded state suggests that without ligand, the L2′ loop is in a conformation that differs from the active state. If the unliganded enzyme "rested" in the active (L2′ up) conformation, as suggested by the mature caspase-3 structure[31], then this mutation should have the same energetic effect on the active-site liganded form as on the unbound form. As this is not the case, our analysis supports the notion that structural rearrangements that are observed in the existing crystal structures of caspase-7[37,39,41,42,61] are the biologically relevant conformations.

In the context of the other mutagenesis data, the Y211A mutation indicates that interactions between the backbones of the L2′ loops covering the allosteric pocket are the

major energetic contributors to the allosterically inhibited/zymogen complex. In the active conformation, the interactions between the L2 loop from one half of the dimer and the L2′ loop from the other half of the dimer have a tremendous stabilizing effect on caspase-7, as the addition of active-site inhibitor increases the apparent thermal stability of the complex by 12–19°C. Thus, the L2′ loop appears to play important roles in both substrate binding and maintaining caspase-7 in an inactive yet still dimeric conformation.

We report a large perturbation (10- to 20-fold) in the K_m of two caspase-7 variants (K212A, I213A) with alanine substitutions in the L2′ loop. Although the L2′ loop is further away in both amino acid sequence and in space than the L2 loop is from the catalytic residues, our reported mutations had as a great an effect on substrate binding and catalytic efficiency as did mutations in the L2 loop. Clark and coworkers[32] mutated caspases-3 residues 167, 169, 173, and 203 (homologous residues in caspase-7 are 190, 192, 196, and 229) to alanine. The residue with the most significantly impaired substrate-binding properties was D169A, in which K_m was increased 35-fold relative to wild type. It is significant that our mutants had the same magnitude effect on substrate binding as mutations in the L2 loop because structurally the two loops also appear to act in concert.

Another region of the caspase substrate-binding loop bundle that has been widely discussed as being critical for caspase activity is the DDD "safety catch" in the L2′ loop in caspase-3. This "safety catch" comprises residues 179–181 (caspase-7 residues 205–207) and was found to be critical for activation of procaspase-3[64]. The DDD "safety catch" region is further from the active site, has not been observed in crystal structures, and is unlikely to make physical buttressing interactions like those we observe in the L2′ buttressing region. Although the DDD region is essential for proper cleavage and

activation of caspase-3, the L2′ loop may be a more tractable target for new methods of caspase regulation.

Before this work, the L2′ loop was implicated in both the active and zymogen/allosterically inhibited conformations. In the active state, L2′ is part of the substrate-binding loop bundle. In the zymogen/allosterically inhibited state, L2′ forms a lid occluding the allosteric cavity. Interconversion between the active and inactive conformations had been observed, but the energetic contributions of L2′ to these two conformations were unclear. Our work demonstrates that residues 212 and 213 are critical for proper functioning of the L2′ loop, particularly in binding substrate. We have shown that I213, in particular, serves as a critical buttress for holding the L2 loop (which is connected to the catalytic cysteine) in place and thus stabilizing the active form of the enzyme. If L2′ is not in the up conformation, the entire loop bundle is destabilized. This buttressing region is present in the homologous conformation in structures of the other active caspases indicating that the results we report may be relevant to the entire caspase family.

This understanding of the role of the L2′ loop as critical for formation of the active-site loop bundle also provides clues toward new mechanisms for pharmacological control of caspases. Molecules that bind overlapping with the buttressing region occupied by I213 and neighbors should work as allosteric inhibitors of caspase function. Similarly, molecules that bind behind the L2/L2′ interacting region [which is on the back side of caspase-7 in (Figure 6 (A), Figure 7 (C), Figure 8 (E)] should stabilize the active conformation and serve as heterologous caspase activators, which are a long-sought entity. Given recent evidence that the uncleaved form of caspase-7 can support catalysis

in a hemicleaved heterodimer,[62,63] this mechanism of activation might even prove more relevant for activation of procaspases.

Materials and Methods

Caspase-7 mutant generation, expression, and purification

Wild-type and mutant versions of caspase-7 were expressed from a two plasmid expression system[39]. The constructs encoding residues 50–198 comprise the large subunits and were contained in the pRSET(Amp^R) plasmid. The constructs encoding residues 199–303 and one codon for Q plus a six-histidine tag are contained in the plasmid pBB75 (Kan^R) plasmid[72]. Mutants in the small subunit were generated using QuikChange (Stratagene) site-directed mutagenesis. The recombinant large and small subunits were coexpressed in *E. coli* in 2× YT media grown for 18 h after induction with 1 m*M* Isopropyl β-d-1-thiogalactopyranoside at an OD_{600} of 0.6. Wild-type and mutant caspase-7 variants were purified using Ni-affinity liquid chromatography (HiTrap Chelating HP, GE). After binding of protein to the affinity column, the protein was eluted with a step gradient from 50 to 250 m*M* imidazole. Protein was diluted to 50 m*M* NaCl and then purified using a Macro-Prep High Q ion exchange column (Bio-Scale Mini 5 mL, Bio-Rad) with a linear gradient from 50 to 500 m*M* NaCl in 20 m*M* Tris buffer pH 8.0, with 2 m*M* DTT. Protein eluted in 120 m*M* NaCl and 20 m*M* Tris pH 8.0 was assessed for purity by SDS-PAGE to be 98%+ pure and stored in elution buffer at −80°C.

Caspase-7 I213A crystallization and X-ray data collection

To prepare I213A crystals, 29 μM I213A protein in a buffer containing 120 mM NaCl and 20 mM Tris pH 8.0 was incubated at room temperature with a 3:1 molar ratio of

DEVD-CHO (asp–glu–val–glu-aldehyde, BioMol) to protein for 3 h. Optimal protein labeling occurred with the addition of enzyme to the small volume of DMSO-solvated inhibitor. The extent of enzyme labeling was confirmed by activity determination. Protein was then concentrated using Millipore Ultrafree 5K NMWL membrane concentrators (Millipore) to 11 mg/mL as assessed by absorbance at 280 nm. Crystal tray setup was conducted on ice with cooled buffers. Crystals were grown in 2 μL hanging drops with mother liquor consisting of 14% polyethylene glycol 3350, 300 mM diammonium hydrogen citrate at pH 5.6, 10 mM guanidine hydrochloride, and 10 mM DTT in a 1:1 ratio of protein mother liquor. Crystals grew to a maximum of 200 μm in 48 h at 4°C. Crystals were cryoprotected in 14% polyethylene glycol 3350, 300 mM diammonium hydrogen citrate at pH 5.6, 21% glycerol, and 10 mM DTT with a 60 second incubation, then frozen by rapid immersion in liquid N_2. Data were collected on an R-axis IV^{++} detector using a Rigaku X-ray generator and showed diffraction data to 2.4 Å. Indexing, integration, and scaling were carried out on HKL2000[73]. Low I/σ in the high-resolution shells forced the dataset to be restricted to 2.61Å. Only 47° of data were collected, resulting in low overall data redundancy, because of a power outage leading to the loss of the crystal.

Structural determination

Phase information was obtained by molecular replacement using 1F1J as the search model with residues 194–196/494–496 and 212–213/512–513 omitted during searches with CCP4 AMoRe[74]. Tyrosine at position 211/511 was not present in the parent model. Removed residues were rebuilt into unambiguous density using O,[75] then refined with three rounds of Refmac restrained refinement using medium strength NCS averaging

restraints. Waters were inserted with multiple rounds of ARP/Warp then checked for stereochemical viability manually. The final refined model contains residues 58–196, 211–303 for chain A and 57–196, 211–304 for chain B. The inhibitor is modeled as chains C and D, with residue numbers 701–705 and 801–805, respectively, and 119 ordered waters as chain W.

Caspase activity assays

Enzyme concentrations were determined by active site titration with suicide substrate DEVD-Fluoromethylketone (BioMol) in caspase activity buffer containing 100 mM Hepes pH 7.0, 10% polyethylene glycol 400, 0.1% CHAPS (3-[(3-cholamidopropyl)dimethylammonio]-1-propanesulfonate), 5 mM β-mercaptoethanol, and 5 mM $CaCl_2$ against flourogenic substrate, DEVD-AMC (7-Amino-4-methylcoumarin), Ex365/Em495 (BioMol). Active site titration setups were incubated over a period of 4 h in 120 mM NaCl, 20 mM Tris pH 8.0 at nanomolar concentrations. Optimal labeling was observed when protein was added to DEVD-FMK solvated in DMSO in 96-well V-bottom plates, sealed with tape, and incubated at room temperature. Ninety microliters of aliquots were transferred to black-well plates in duplicate and assayed with 50-fold molar excess of substrate. For kinetic measurements, 50 nM protein (500 nM in the case of K212A, I213A) was assayed over 0–180 μ*M* DEVD-AMC (0–500 μM for K212A, I213A) over a course of 7 min. Assays were performed at 37°C in 100 μL volumes in microplate format using a Molecular Devices Spectramax M5 spectrophotometer. Initial velocities versus substrate concentration were fit to a rectangular hyperbola using GraFit software (Erithacus Software) to determine kinetic parameters $K_m$and k_{cat}.

DICA inhibition and reduction assays

DICA removal assays were conducted on both WT and P214A proteins buffer exchanged into 10mM sodium phosphate buffer pH 7.5. Proteins were then incubated for 1h in 10X molar excess of DICA, followed by a buffer exchange to remove excess DICA. Proteins were then diluted to 110 nM as assayed by A280, and incubated in a 2 fold dilution series of DTT (starting at 100 μM) for 30 minutes and assayed for activity against a control under identical conditions without DICA.

Sizing column analysis

The oligomeric state of the caspase-7 WT and P214A in the presence and absence of DEVD-CHO was determined by running protein samples on a Superdex 200 10/300 GL (GE Healthcare) gel-filtration column. Both inhibited and apo proteins were then buffer exchanged into 20mM Tris, pH 8.0, 100mM NaCl using Millipore Ultrafree 5K NMWL membrane concentrators (Millipore), and 200 μl of protein (1.2 – 2 mg/ml) was loaded onto the column. Protein was eluted with 20mM Tris, pH 8.0, 100mM NaCl. The identity of the peak fractions and the loaded protein sample were analyzed by SDS-PAGE. Four different molecular weight standards (Albumin, Chymotrypsinogen A, Ribonuclease A, Ovalbumin) from the gel-filtration calibration kit (GE Healthcare) were run in the same conditions and a standard plot was generated to calculate molecular weights of caspase-7 WT and P214A in presence and absence of inhibitor.

Analytical ultracentrifugation

Analytical ultracentrifugation of caspase-7 P214A in equilibrium experimental mode were run on Beckman Optima XL-1 analytical ultracentrifuge using an An-60 Ti rotor with a 6 cell equilibrium cuvette (Spin Analytical). Proteins were buffer exchanged into

10mM sodium phosphate buffer pH 7.5 and concentrated using Millipore Ultrafree 5K NMWL membrane concentrators (Millipore). Protein concentrations of 15 μM, 10 μM and 1 μM were run to equilibrium at 14,000, 18,000 and 21,000 RPM for 14 hours each at 20°C and monitored by UV at 280 nM in columns heights of ~2.8mm. Density of the buffer and the specific volume of the protein were calculated with ULTRASCAN II[76], while data for the 15 μM and 10 μM were analyzed by SEDFIT[77] and SEDPHAT[77,78] using the continuous distribution model. Additional analytical ultracentrifugation data analysis support was graciously supplied by Dr. Jeffery Lary of the University of Connecticut.

Thermal stability determination

Thermal denaturation of wild-type or caspase-7 variants was monitored by loss of circular dichroism signal at 222 nm on a J-715 circular dichroism spectrometer (Jasco) with a PTC-348WI Peltier controller. Inhibited protein was prepared by 4-h incubation with 3M equivalents of DEVD-CHO (asp–glu–val–glu-aldehyde, BioMol) at a concentration of 6–12 μM in 120 mM NaCl, 20 mM Tris pH 8.0. Both inhibited and apo proteins were then buffer exchanged into 10 mM phosphate buffer, pH 7.4 using Millipore Ultrafree 5K NMWL membrane concentrators (Millipore) for repeated concentration and dilution until NaCl concentration was below 1 nM and protein concentration of ~12 μM as assessed by absorbance at 280 nm. Data were collected in duplicate on separate days and fit using Origin Software (OriginLab) using sigmoid fit.

Assessment of generated cavity volumes

Models of the alanine-mutant caspase-7 variants were generated in PyMol[79] and analyzed with Pocketfinder[80] to predict the volume of pockets created by alanine mutations.

CHAPTER III

A DESIGNED REDOX-CONTROLLED CASPASE

Abstract

Caspases are a powerful class of cysteine proteases. Introduction of activated caspases in healthy or cancerous cells results in induction of apoptotic cell death. We have designed and characterized a version of caspase-7 that can be inactivated under oxidizing conditions like those found extracellularly, then reactivated under reducing, intracellular conditions. This version of caspase-7 is allosterically inactivated when two of the substrate-binding loops are locked together via an engineered disulfide. When this disulfide is reduced, the protein regains full function. The inactive loop-locked version of caspase-7 can be readily observed by immunoblotting and by mass spectrometry. The reduced and reactivated form of the enzyme observed crystallographically is the first caspase-7 structure in which the substrate-binding groove is properly ordered even in the absence of an active-site ligand. In the reactivated structure, the catalytic dyad cysteine-histidine are positioned 3.5 Å apart in an orientation that is capable of supporting catalysis. This redox-controlled version of caspase-7 is particularly well-suited for targeted cell death in concert with redox-triggered delivery vehicles.

Introduction

The loops that form the caspase substrate-binding groove are extraordinarily mobile and controlling their conformation controls caspase function. Amongst caspases, the most thorough molecular descriptions of activation, regulation and inhibition have been in

caspase-7. In the procaspase-7 zymogen, the active-site loops are disordered and incompatible with substrate binding[39]. Upon activation the active-site loops are partially ordered, but the L2' loop remains in an inactive, down conformation (Figure 18 A). From this state caspase-7 is competent to bind substrate or to bind allosteric inhibitors, mutually exclusively. When substrate binds, L2' moves to an up conformation, forming a foundation or buttress beneath the L2 loops[37,39,41,61]. When caspase-7 binds allosteric inhibitors in a cavity at the dimer interface, the L2' loops are held in the down conformation and the enzyme is inactive[81,82]. Thus controlling caspase-7 loop conformations is an effective way to modulate activity.

Caspases are found only intracellularly and must function inside the cell to activate apoptosis. Caspases make use of a cysteine-histidine active-site dyad for proteolysis. The catalytic cysteine nucleophile must be maintained in a reduced state to function. Glutathione is the most prevalent cellular redox component. Intracellular glutathione concentrations in healthy cells range from 1-5 mM, with cancerous cells showing concentrations up to 7-fold higher,[83] whereas extracellular concentrations are 1-4 μM[84]. Thus differences in redox environment provide an intriguing means of differentially controlling intracellular and extracellular activity.

Caspases are exquisitely useful cellular players. Once activated they singlehandedly induce the demise of a cell. There is a great deal of interest in small-molecule activation of caspases, however this has been challenging to achieve[85-87]. Another option for selective killing of cells therapeutically is the use of pre-activated caspases. Injecting activated caspases, or any activated protease, into circulation is likely to produce many unwanted and deleterious effects. In order to avoid these complications while still taking

advantage of the cell-killing properties of activated caspases, we aim to engineer a version of the executioner caspase-7 that can be held in an inactive conformation in the oxidizing extracellular environment, but regains full activity in reducing intracellular conditions. Targeted delivery of this type of engineered caspase could potentially provide a safe and effective mechanism of selectively killing unwanted cell types.

Results

Rational design of a redox activatable caspase-7.

The aim of this work is to control the conformation of the L2' loops in order to allosterically regulate caspase-7 function. In the crystal structure of caspase-7 bound to allosteric inhibitors FICA or DICA, caspase-7 attains a conformation that is distinct from the mature or substrate-bound forms. We have previously shown that caspase-7 likewise exists in two distinct conformations in solution[88], a bound active state and an unbound inactive state. In the inactive state with no substrate present, the L2' loops rest over the allosteric cavity and do not contribute to the formation of the active-site loop bundle. The L2' loops from the opposite sides of the dimer form an anti-parallel structure with a nearly 90° crossing angle. The residues with the closest approach are the two R210 residues, with a Cα-Cα distance of 5.5Å (Figure 18 A).

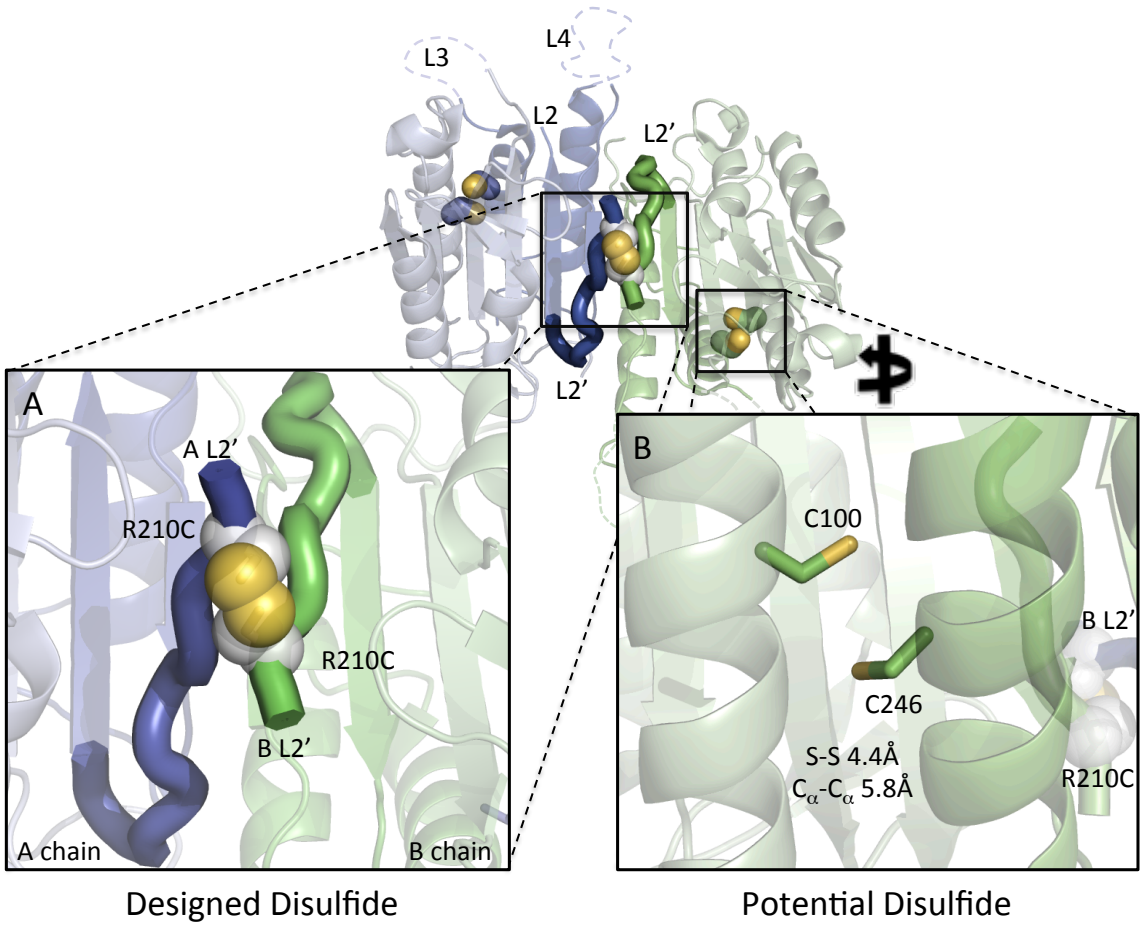


Figure 18: Design of a reductant activatable caspase-7

The allosterically inactivated caspase-7 dimer consists of two chains (A chain, green; B chain, blue). (A) An introduced disulfide between R210C residues on the two chains (white and yellow spheres) is designed to lock the protein into the inactive conformation by keeping the L2' loops in the down conformation. Formation of this disulfide would result in a covalently crosslinked small-subunit homodimeric species. (B) Pairs of native cysteines (C100 and C246, blue and yellow sticks) are computationally predicted to form disulfides on each side of the dimer. These disulfides would drive formation of a large subunit-small subunit crosslinked species.

We modeled pairwise mutations to cysteine for all residues in the L2' loops then used the MODIP[89] server to score the probability of forming various disulfides. MODIP scored an R210C-R210C disulfide in the C category on a scale of A (most favorable) to F (least favorable). This intermediate score may be related to the fact that all Cα positions are fixed during MODIP modeling. The L2' loop is necessarily flexible, as it undergoes significant rearrangement during the caspase lifecycle, thus a score of C may still indicate

a favorable interaction. In contrast, a second potential native disulfide, which we subsequently observed crystallographically (Figure 19), was scored as an A.

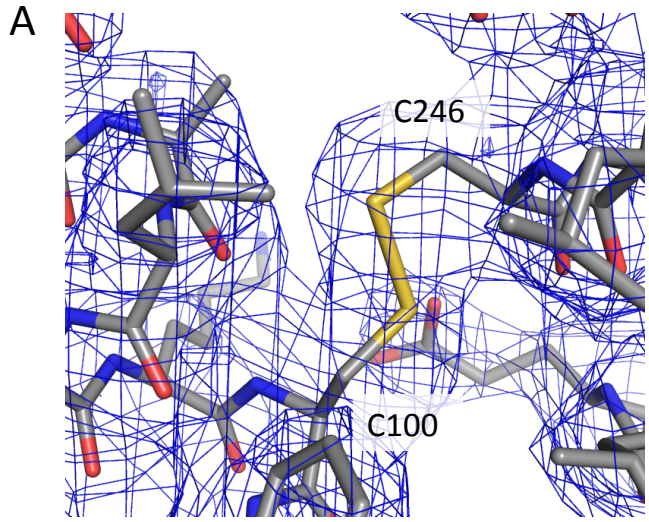


Figure 19: Observation of a C100 and C246 disulfide.

(A) A C100-C246 disulfide is visible between the large and small subunits in a crystal structure of caspase-7 with an uncleavable peptide inhibitor bound in the active site. $2F_o$-F_c density contoured at 1.0 σ is shown. (B) Sequence alignment of caspases in the region of C100 and C246. The potential disulfide linking the large and small subunits is unique to caspase-7.

Our first design attempt was the substitution R210C. Purification of the single R210C variant was not successful, due to low protein yields. We speculated that the formation of mixed disulfides may have contributed to low protein yields.

We found that removal of one native cysteine, which improved protein yields and eased analysis was essential. Caspase-7 is unique among caspases in that it contains a potential disulfide interaction between residues 100 (large subunit) and 246 (small subunit) (Figure 18 B). We engineered out the native C100-C246 disulfide by making the C246S substitution. C246S had similar catalytic parameters to wild-type (Table 1). We combined this with the disulfide locking R210C substitution. The doubly substituted variant (R210C/C246S) had good catalytic activity and exhibited K_M and k_{cat} very similar to wild-type (Table 3, Figure 20).

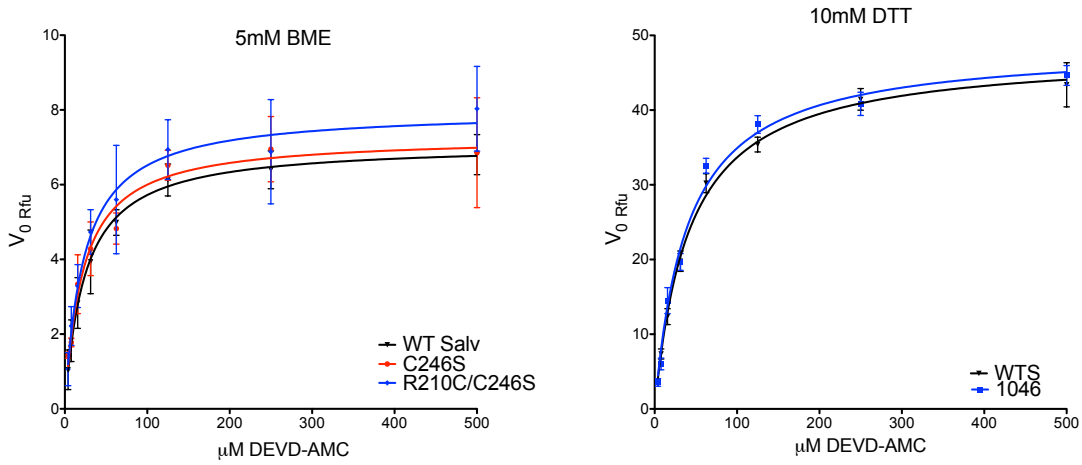


Figure 20: Kinetics curves for caspase-7 WT, C246S, R210C/C246S

Raw RFU (relative fluorescence unit) data for substrate titration of caspase-7 WT, C246S, and R210C/C246S in both 5 mM BME and 10 mM DTT. Michaelis-Menten curve fitting was applied to determine k_{cat} and K_m with protein concentration held constant.

	5 mM BME			10mM DTT	
	WT	C246S	R210C C246S	WT	R210C C246S
k_{cat} (s^{-1})	0.36 ± 0.01	0.33 ± 0.01	0.40 ± 0.01	2.42 ± 0.07	2.46 ± 0.06
K_m (μM)	23.98 ± 2.02	21.46 ± 2.77	22.65 ± 2.85	41.87 ± 4.15	39.36 ± 3.42

Kinetic parameters are based on substrate titrations measured from independent triplicate dilutions of substrate on two days.

Table 3: Kinetic parameters for caspase-7 WT, C246S, R210C/C246S

Measured kinetic parameters for WT, C246S, and R210C/C246S do not show significant variances, suggesting the mutations do not have an effect on enzyme kinetics. Similarly, while overall enzymatic rate increases in the presence of higher concentrations of reductant, the increase is statistically identical between WT and R210C/C246S.

The observed differences between 5 mM and 10 mM DTT can be attributed to the fact that DTT has a higher reduction potential. In cysteine-histidine proteases an oxidized cysteine would disable the catalytic activity, as mechanism is dependent on the histidine abstracting a proton creating a reactive thiolate. The stronger reductant present eases the formation of this thiolate species and helps to facilitate the reversal of the tetrahedral covalent intermediate by H_2O increasing the apparent activity of caspase-7.

Observation of the designed disulfide dimer.

The variant R210C/C246S is designed such that mildly oxidizing conditions should drive formation of a disulfide, locking the two L2' loops together across the dimer interface and allosterically inhibiting caspase-7. Because the L2' loops are the *N*-termini of the small subunits, formation of the inactivating R210C-R210C disulfide will generate covalent small subunit dimers. To measure formation of small-subunit dimers, we exposed wild-type or R210C/C246S caspase-7 to native (reducing conditions) or mildly oxidizing conditions (300 μM cystamine). The bands on a denaturing acrylamide gel were unambiguously identified by a combination of mass spectrometry and

imunoblotting with antibodies specific for the large or small subunits of caspase-7 (Figure 21, Figure 22).

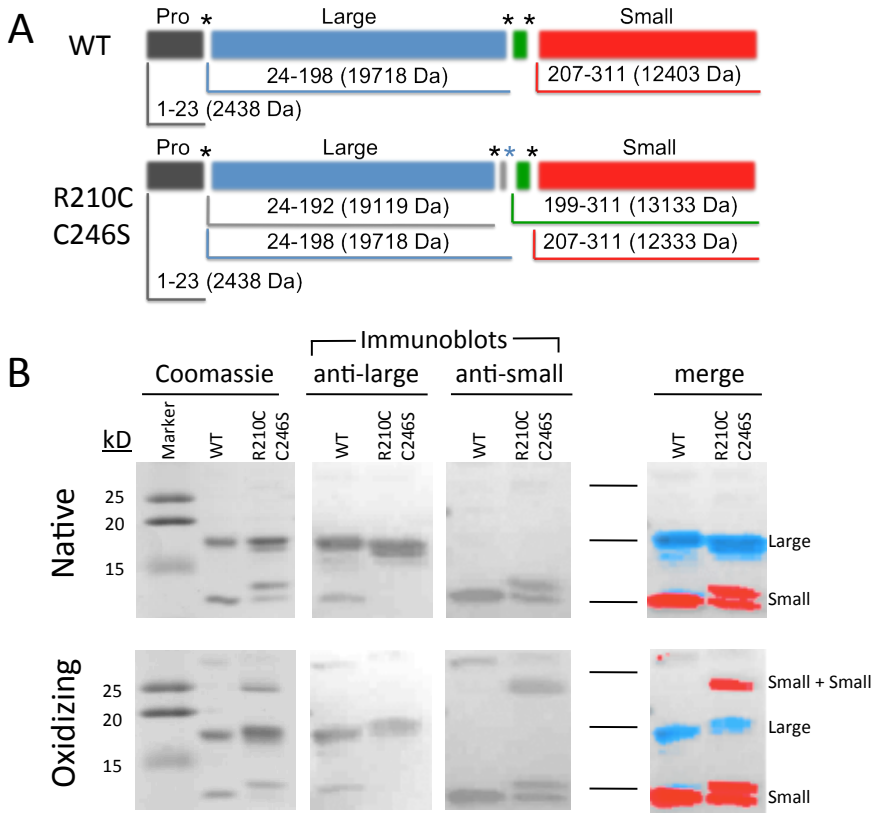


Figure 21: Incubation of caspase-7 R210C/C246S under oxidizing condition results in formation of a unique disulfide between the two small subunits of caspase-7.

(A) Domain structure and molecular weights of WT or R210C caspase-7. (B) WT and R210C/C246S caspase-7 were incubated in native conditions (2 mM DTT) or in oxidizing conditions (300 µM cystamine) then analyzed on a coomassie-stained SDS gel or by western blot with a caspase-7 anti-large subunit or anti-small subunit antibody. The three gels were merged with the anti-large subunit signal colored blue and anti-small subunit signal colored red. The two small subunit bands in the native conditions are alternatively cleaved versions of caspase-7, with cleavages at naturally occurring residues 199 or 207.

Caspase-7 has two major auto-activating cleavage sites in the intersubunit linker at residues 198 and 206 (Figure 21 A). In wild-type caspase-7 the intersubunit linker (residues 199-206) is removed so the prodomain-deleted large subunit is composed of residues 24-198 and the small subunit consists of residues 207-311. For our designs, we used a construct that allows constitutive expression of the two-chain (active) form of

caspase-7. R210C/C246S caspase-7 was produced as two polypeptides of residues 1-198 and 199-311. In addition to removal of the prodomain, the large subunit was also processed at a minor cleavage site (residue 192, which has been extensively probed by Denault and co-workers[90]) generating some large subunits composed of residues 24-192. The small subunit of R210C/C246S appears in two forms (Figure 21 A,B), both with the intersubunit linker (residues 199-311) and without the intersubunit linker (residues 207-311). Under oxidizing conditions the C100-C246 disulfide can be observed in wild-type caspase-7 as a large-small dimer as a minor product. Importantly, the expected small-subunit (small-small) dimer band is strongly formed in R210C/C246S, but is not observed in wild-type caspase-7. This demonstrates that the designed dimer forms as expected but spurious and random disulfides do not form. The small-small dimer could also be readily observed by mass spectrometry (Figure 22, Table 4).

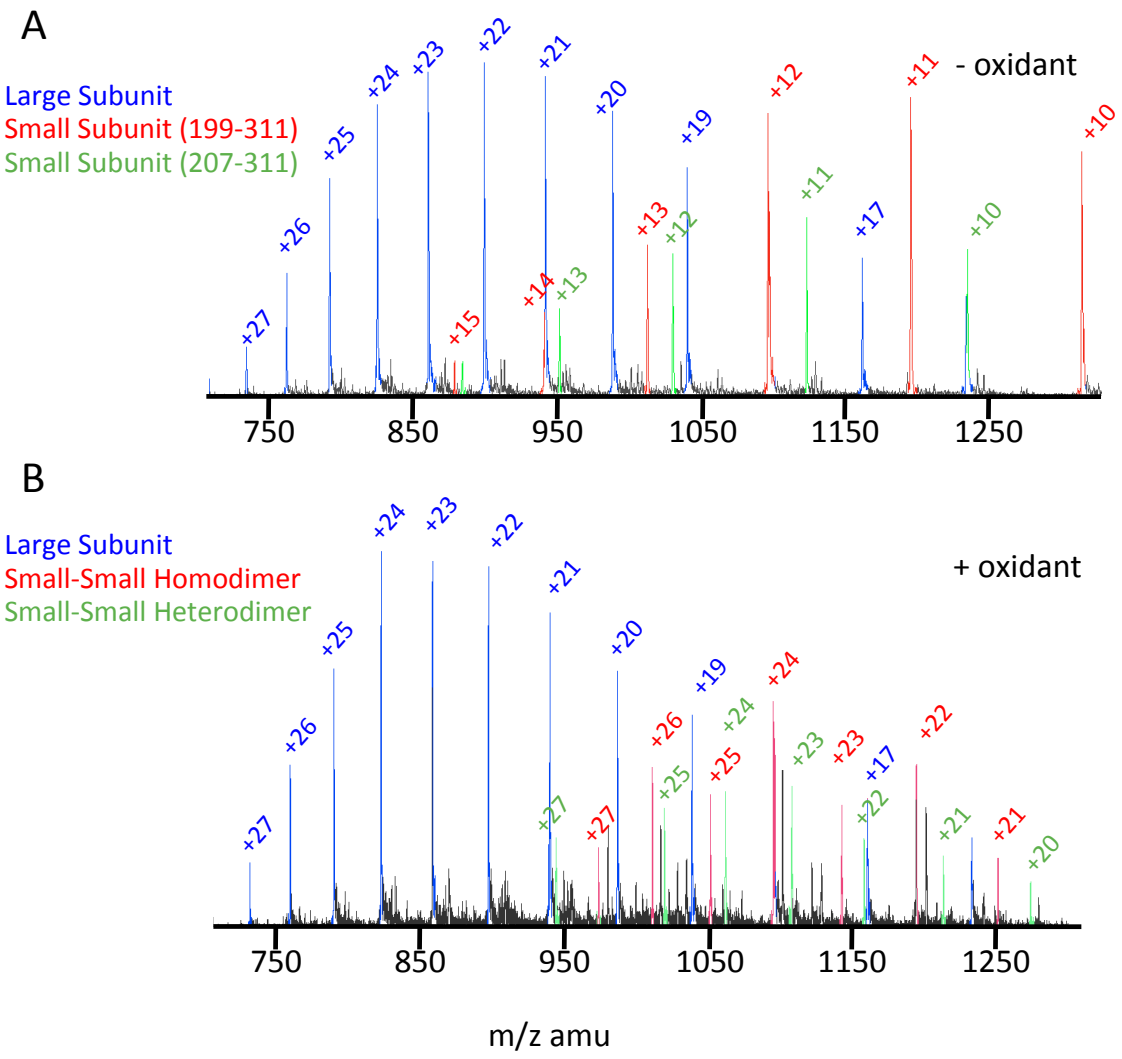


Figure 22: The small-small subunit dimer can be observed by mass spectrometry

R210C/C246S was incubated in the presence or absence of oxidant (1.5 mM cystamine) then analyzed by mass spectrometry. (A) In the absence of oxidant clear ions for the large subunit (blue) and the two cleavage variants of the small subunit (red, residues 199-311; green, 207-311) were observed. (B) In the presence of oxidant, identical ions were observed for the large subunit (blue). For the small subunits twice as many ions were observed indicating a doubling of the molecular weight due to dimerization of the 199-311 species (red). A hybrid small subunit dimer between the 199-311and the 207-311 species (green) is also observed.

No Oxidant			Oxidant		
Mass (avg.)	m/z	Charge (z)	Mass (avg.)	m/z	Charge (z)
19718.0			19718.7		
	705.2	28		705.3	28
	731.3	27		731.3	27
	759.4	26		759.4	26
	789.7	25		789.8	25
	822.6	24		822.6	24
	858.3	23		858.3	23
	897.3	22		897.3	22
	940.0	21		940.0	21
	986.9	20		986.9	20
	1038.8	19		1038.8	19
	1096.4	18		1096.5	18
	1160.9	17		1160.9	17
	1233.4	16		1233.5	16
	1315.4	15		1315.5	15
	1409.5	14		1409.5	14
	1517.8	13		1517.8	13
	1643.9	12			
	1793.6	11			
13133.1			26265.7		
	730.6	18		730.6	36
	773.5	17		751.6	35
	821.8	16		773.4	34
	876.6	15		796.9	33
	939.1	14		821.8	32
	1011.3	13		848.3	31
	1095.4	12		876.5	30
	1194.9	11		906.7	29
	1314.3	10		939.1	28
	1460.2	9		973.8	27
	1642.7	8		1011.2	26
				1051.6	25
				1095.4	24
				1143.0	23
				1194.9	22
				1251.8	21
				1314.3	20

Table 4: Tabulated MS deconvolution ion data of R210C/C246S

Tabulated deconvoluted ions from the R210C/C246S mass spectra collected in the presence and absence of oxidant. Observed protein masses and corresponding mass to charge (m/z) states are listed. All deconvolutions were gated for masses between 10kD and 30kD.

Under reducing conditions only monomeric large and small subunits of R210C/C246S were observed. Under oxidizing conditions, small-small dimers were readily observed, further confirming the success of the design. Interestingly, both homo-dimers and hetero-dimers of the two small subunit cleavage variants are observed. For this design to be fully successful it is essential that small-small dimers can strongly form, inactivating R210C/C246S, and subsequently be reduced, resulting in the return of full activity.

Inactivating disulfide forms reversibly

R210C/C246S can be inhibited under mildly-oxidizing conditions (Figure 23) in a manner that is consistent with observation of the small subunit dimers (Figure 21, Figure 22).

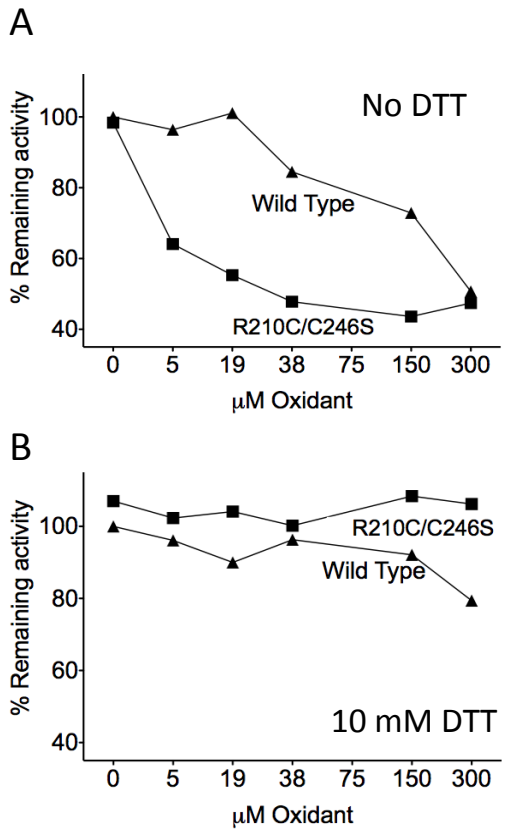


Figure 23: Disulfide bond formation induces the allosterically inactive state of caspase-7, which can be reactivated by the presence of reductant.

Caspase-7 proteolytic activity was monitored in the presence of increasing concentrations of oxidant (cystamine). (A) Oxidation of the R210C is capable of inactivating R210C/C246S at much lower concentrations than the active site can be oxidized in wild-type caspase-7 (WT). (B) In the presence of reductant (10 mM DTT) full activity is recovered. Data is based on activity assays measured from independent duplicate data sets on two separate days.

Inactivation of oxidized R210C/C246S can be monitored either directly by assaying cleavage of a fluorescence substrate (Figure 23), or indirectly by observing that cleavage at the 192 site is slowed only under oxidizing conditions (Figure 21). We tested the reversibility of R210C/C246S inhibition by incubation in reductant at a roughly intracellular redox potential. R210C/C246S was able to fully recover from oxidation-induced inhibition. Inhibition of wild-type caspase-7 is likely the result of mixed disulfide formation between the catalytic cysteine and the cystamine oxidant, which has previously been reported[91] and is also fully reversible under these conditions.

We observed caspase-7 R210C/C246S in an active conformation in the crystal structure. There are two possible explanations for why the oxidized form of the enzyme was not observed. One is that the latent presence of month old DTT (2mM) from the purification was not sufficiently overwhelmed by 1.5 mM oxidant, as was seen in the mass spectroscopy experiment. This would have resulted in the reduction of the disulfide during crystallization, though we consider this unlikely. A second possibility is that the disulfide was reduced by synchrotron radiation, and we discuss that possibility here.

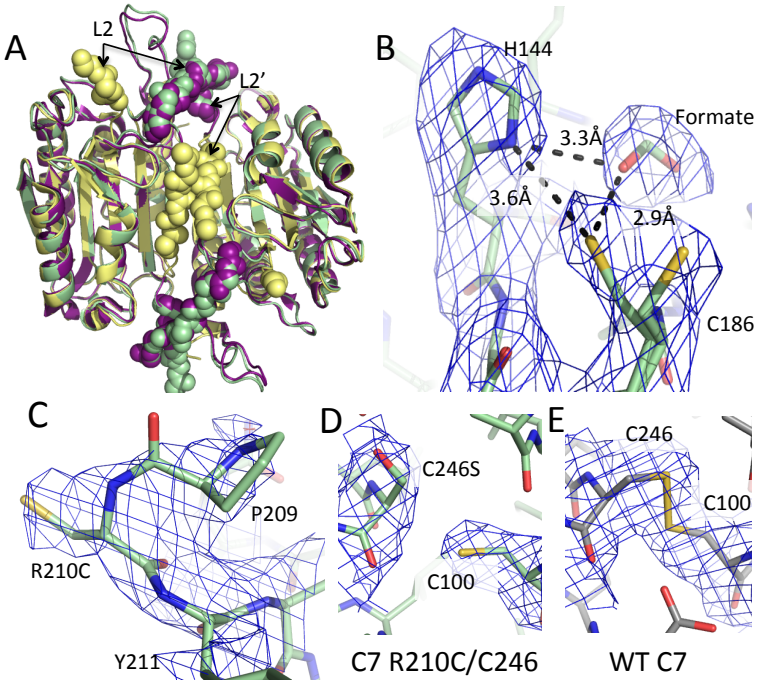


Figure 24: Caspase-7 R210C/C246S can reversibly attain the active conformation.

(A) Oxidized R210C/C246S was crystallized but was reduced prior to or during data collection. Reduced caspase-7 R210C/C246S (3R5K, green) attains a conformation like that of WT caspase-7 bound to substrate (1F1J, purple) with L2 and L2' in the up conformation interacting across the dimer interface. This is in contrast to the allosterically inhibited conformation of caspase-7 (1SHJ, yellow) in which the L2' loop is folded over the allosteric site and the two L2' loops make contact with one another. (B) The catalytic residues H144 and C186 are in a conformation that is close to ideal for catalysis. Additional $2F_o$-F_c density contoured at 1 σ density was observed in the S1 pocket was modeled as a formate molecule from the crystallization solution. The density around the catalytic cysteine (C186) was larger than one cysteine residue. It was best modeled as two cysteine conformations each with 50% occupancy. (C) $2F_o$-F_c density contoured at 0.8 σ for R210C. (D, E) $2F_o$-F_c density contoured at 1 σ for R210C/C246S (D) and wild-type (E) caspase-7

We solved the structure of R210C/C246S in two independent crystals (Table 5). In the structures from both of the crystals it was clear that the R210C disulfide had been reduced. We were pleased to see that as was the case for chemical reductants, the R210C-R210C disulfide could be reversed in the crystal by ionizing radiation, allowing the L2' loop to adopt the active conformation (Figure 24 A). This was not surprising, as ionizing radiation has been reported to reduce disulfides[92]. The radiation applied to the crystal at the APS synchrotron x-ray source has a flux on the order of 10^{19} photons/sec/mm^2. It has been reported that sources with flux of 10^{15} photons/sec/mm^2 can induce disulfide reduction and global protein relaxation[93]. In an attempt to observe the disulfide-containing form of R210C/C246S we limited x-ray exposure to one second with the APS 24-ID-C x-ray beam attenuated to 7% and solved the structure using the minimal data necessary (30 degrees of data, 20 frames, 1 sec/frame, 83% completeness). Even with only 20 seconds total exposure to the crystal, we were unable to observe any oxidized disulfide, suggesting that the inactivating disulfide is quite labile. We can roughly estimate the lability of the R210C disulfide by comparison to other caspase-7 disulfides that do withstand x-ray interrogation. DICA and FICA, which bind to caspase-7 via disulfides with C290, withstand x-ray radiation and are visible in the crystal structure[42]. The disulfide linking the large and small subunits of caspase-7 (C100-C246) also withstands x-ray radiation under certain conditions (Figure 19). Thus of the three types of disulfides observed to date in caspase-7, the R210C disulfide may be the most labile. All caspase-7 structures observed to date show the same crystal packing arrangement. Regardless of the loop conformation, the active-site loops always point toward the large solvent channels. Thus it is not surprising that the loop reorganization we observed

between the oxidized, disulfide-locked R210C/C246S and the reduced, unlocked form can be accommodated and interconverted within the same crystal.

Diffraction Data	Native Data set 1	Native Data set 2
Wavelength (Å)	0.9792	0.9792
Resolution Range (Å)	48-2.86	48-3.1
Oscillation Range per frame	1.0°	1.5°
Measured reflection (*n*)	106848	28496
Unique reflections	20763	17353
Completeness (%)	99.2 (97.2)	83.6 (87.9)
Redundancy	5.2 (4.7)	1.6 (1.6)
$\langle I/\sigma I \rangle$	16.0 (2.7)	16.4 (2.0)
Space Group	$P3_221$	$P3_221$
a = b (Å)	89.85	89.98
c (Å)	185.91	185.11
α = β (°)	90	90
γ (°)	120	120
R_{sym}	12.7 (69.6)	9.1 (69.4)
Refinement Statistics		
Atoms (n)	3900	
Water molecules	103	
R_{work} (%)	18.81	
R_{free} (%)	25.13	
RMSD bond length (Å)	0.010	
RMSD bond angle (°)	1.197	
Average B-factor (Å^2)	58.93	

Table 5: Crystallographic and Refinement parameters of R210C/C246S

Data in parenthesis are for the highest resolution shell

Catalytic site geometry

The reduced form of R210C/C246S is in a conformation that is primed for substrate binding (Figure 24). Typically, in the absence of an active site ligand, the L2' loop is in a catalytically inactive conformation[39] similar to the allosterically inhibited form[42] (Figure 24 A, yellow spheres). In our structure the presence of formate at high concentrations (2.1M) in the crystallization buffer is sufficient to nucleate formation of the substrate-

binding conformation (Figure 24 A, green and purple spheres). Here formate serves as a surrogate for the aspartic acid that normally occupies the S1 pocket (Figure 24 B). Because structures of caspases in the active conformation have always been solved in the presence of the covalent inhibitor, a structure of WT caspase-7 with formate in the active is not available. Based on evidence in which a similar phenomenon has been observed for caspase-1 where malonate assists in refolding caspase-1 and serves as a mimic for aspartic acid bound in the active site[94]. Although the structure of wild-type caspase-7 has not been determined in formate-containing conditions, we believe that WT protein would exhibit a similar substrate-binding groove structure, nucleated by formate, in the presence of high concentrations of formate.

In R210C/C246S the catalytic dyad (H144 and C186) are in a closer to optimal arrangement for catalysis (3.6Å) that has not been observed before in caspase-7. Papain, another cysteine protease, has been reported to have a cysteine-histidine distance of 3.2Å in a high resolution crystal structure, while computational methods have determined 3.0-3.3 Å to be the optimal distance for catalysis[95,96]. All previous caspase-7 structures in which the substrate-binding loops are ordered relied on covalent modification of the catalytic cysteine[37-39,41,42,61,88] and therefore had H144-C186 distances that were too great (typically ~5.5Å) to support necessary proton abstraction. The density for the catalytic thiol is not unambiguous. We have modeled this density in several ways including positioning a water in one lobe of the density. The unfavorable angle of the hydrogen bonds for water suggested to us that this is not a water molecule. Therefore, we concluded that two conformations of the catalytic thiol may be a more accurate representation. We have thus modeled the density as a 50% mixture of two thiol

conformations (Figure 24 B) though the thiol position nearest the histidine is 35° from an ideal rotamer. R210C is near the *N*-terminus of the small subunit. In caspase-7, it is typically only possible to observe residues starting at residue 212, due to disorder of the *N*-termini. It is therefore not a surprise that the density is weak, nevertheless we were able to visualize and build R210C (Figure 24 C). It was also possible to visualize the substitution in our design, C246S (Figure 24 D). C246S is in a different conformation than C246 locked into a disulfide by an uncleavable peptide (Figure 24 E).

Discussion

The disulfide bond is a useful protein-engineering tool since it can form covalent, redox-controllable bonds at discrete locations from naturally occurring amino acids. Since the advent of site-directed mutagenesis, introduction of disulfide bonds has been a well-used tool for controlling protein conformation. Engineered disulfides have been used to examine functional mechanisms[97,98], stabilize proteins[99,100], understand allosteric coupling[101,102] and even to trap transition states[103]. Redox reversible switches have been developed in a number of proteins including measles virus F protein[104], malate dehydrogenase[105], integrins[106], kinesin[107] and the OxyR transcription factor[108], among many others. We sought to build upon this rich history and design allosterically controlled caspases, a class of proteins capable of regulating life and death of cells.

The criteria that guided the rational design of redox activatable caspase-7 included: (1) motion of the L2' loops should be restricted by disulfide-bond formation; (2) the catalytic efficiency of the untreated protein should mirror that of the wild-type enzyme; (3) full oxidation of the disulfide should result in full inhibition of caspase-7 under extracellular redox environments and (4) introduction into intracellular reducing conditions should

result in full recovery of proteolytic activity. Caspase-7 is a dynamic protein in which loop arrangement plays a key role in activation of the enzyme. In our designed variant R210C/C246S, we placed the inactivation switch at the heart of a critical, active-site loop bundle. In the down confirmation, with the L2' loop pinned over the allosteric cavity, the substrate binding groove is not formed so the protein is unable to bind substrate or catalyze the peptide cleavage reaction. By placing the disulfide over the allosteric site at the intersection of the L2' loops, we can enforce a naturally occurring inactive state using native-like redox conditions. In R210C/C246S we fulfilled all of the design criteria. In vitro activity results and the observation that oxidation of the R210C disulfide could also prevent self-processing at the minor D192 processing site both suggest ready inhibition by disulfide formation. The crystal structure of R210C/C246S that was first oxidized and then *re*-reduced in the crystal demonstrates the successful reversibility of our design. Finally the observation of small-subunit dimers points to the inhibition mechanism by covalent locking of the L2' loops.

Engineering in redox switches in a critical region like the active-site loops bundle could be a risky proposition. We had previously interrogated residues 211-215 in the L2' loop and found that they were amenable to substitution[88]. By extension, we hoped that R210 would likewise be amenable to substitution without negative functional effects. Fortunately, this was the case. All caspases have mobile active-site loops. Although this endeavor used caspase-7 as a template, the approach of pinning the L2' loops should be applicable for any caspase since the active site loops of all caspases are mobile prior to the binding of substrate. Residue 210 is not highly conserved so this position should be amenable to this substitution across the family. Intriguingly, residue 210 is natively a

cysteine in caspase-3, begging the question of whether caspase-3 uses formation of this potential disulfide as part of its unique regulatory mechanisms. In a very high resolution structure of caspase-3, the L2' loop is seen to wrap around the opposite half of the dimer with C210 attaining two solvent accessible conformations[33]. Because caspase-3 has not been crystallized in the absence of a substrate-mimic nor in the allosterically inhibited conformation, the best model of what this region might look like is the allosterically inhibited structures of caspase-7 where the 210 residues make the closest approach. Thus it may be possible to make a loop-locked version of caspase-3. It is intriguing to envision controlling the activity of an engineered, constitutively active and uninhibitable version of the caspase-3 zymogen[34] using our loop-locking design. Thus there are many possibilities for designing loop-locked caspases. Induction of apoptosis by treatment with a variety of allosterically loop-locked caspases should therefore be possible given appropriate means of delivery.

Breakthroughs in targeted nanoparticle and polymer design have begun to make delivery of a much wider range of protein-based drugs feasible (for review[109]). Active targeting of nanoparticles, polymers, dendrimers, and nanogels can be achieved using disulfide functionalization, with the prospect of delivering encapsulated payloads including both small molecules and proteins. The combination of caspase-7 R210C/C246S with a redox-triggered delivery vehicle is therefore an intriguing prospect. If R210C/C246S were encapsulated in a redox-activable nanocarrier, in the relatively oxidizing extracellular environment, R210C/C246S should remain inactive and inaccessibly encapsulated. Upon introduction into the reducing cytosolic environment, the redox sensitive delivery device would release R210C/C246S (Figure 25). Liberated R210C/C246S could then likewise

be reduced, regain function and activate apoptosis, leading to cell death. This scheme is particularly promising in cancers and proliferative disorders, where suppression of apoptosis is nearly always a contributor to the evasion of cell death pathways.

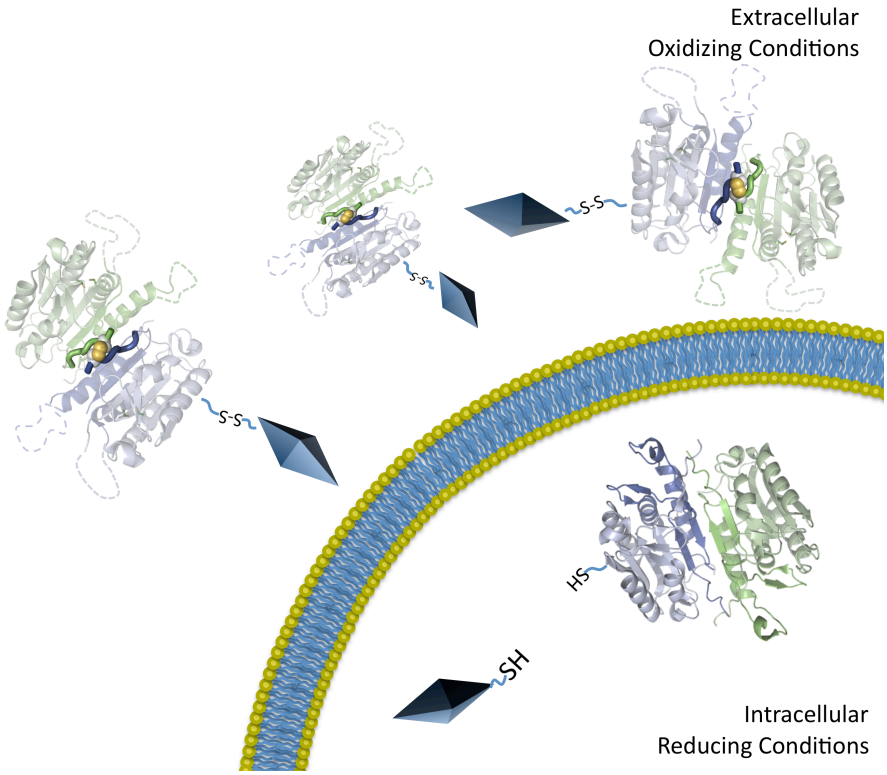


Figure 25: Delivery scheme for caspase-7 R210C/C246S

Capase-7 R210C/C246S is amenable for use with delivery vehicles (diamonds). We have designed the R210C/C246S version of caspase-7 so that under oxidizing extracellular conditions caspase-7 is inactive. Upon entry into the cell, reducing intracellular conditions break the R210C disulfide yielding active caspase-7, which can cleave intracellular targets leading to apoptotic cell death.

Materials and methods

Caspase-7 mutant generation, expression and purification

Wild-type caspase-7 was expressed from a construct encoding residues 1-303 plus codons for the two amino acids LE, then a six-Histidine tag contained in the plasmid pET23-b (Amp^R) plasmid[110]. Expression of caspase-7 variant R210C/C246S was carried

out in a constitutive two chain expression system[27]. Residues 1-198 were followed by a TAA stop codon, a second ribosome binding sequence, and residues 199-303, plus codons for the two amino acids LE and a six-Hisidine tag contained in the plasmid pET23-b (Amp^R). Variants in the protein were generated using QuikChange (Stratagene) site-directed mutagenesis. The recombinant protein was expressed in *E. coli* in 2xYT media grown for 18 hours after induction with 1 mM Isopropyl β-D-1-thiogalactopyranoside at an OD_{600} of 0.6.[111] Wild-type and mutant caspase-7 variants were purified using Ni-affinity liquid chromatography (HiTrap Chelating HP, GE). After binding of protein to the affinity column, the protein was eluted with a step gradient from 2 mM imidazole to 250 mM imidazole. Protein was diluted to 50mM NaCl and then purified using a Macro-Prep High Q ion exchange column (Bio-Scale Mini 5mL, Bio-Rad) with a linear gradient from 50 mM to 500 mM NaCl in 20 mM Tris buffer pH 8.0, with 2 mM DTT. Protein eluted in 120 mM NaCl and 20 mM Tris pH 8.0 was assessed for purity by SDS-PAGE to be greater than 98% pure, and stored in elution buffer at -80°C.

Caspase activity assays

Enzyme concentrations were determined by active site titration with covalent substrate DEVD-CHO (*N*-Acetyl-Asp-Glu-Val-Asp-Aldehyde, Enzo Lifesciences) in caspase activity buffer containing 100 mM Hepes pH 7.0, 10% polyethylene glycol 400, 0.1% CHAPS, 5 mM β-mercaptoethanol, and 5 mM $CaCl_2$ using the flourogenic substrate, DEVD-AMC (*N*-acetyl-Asp-Glu-Val-Asp (7-amino-4-methylcoumarin), Enzo Lifesciences), Ex365/Em495. Active site titration samples were incubated over a period of 2 hours in 120 mM NaCl 20 mM Tris pH 8.0 at nanomolar concentrations. Optimal

labeling was observed when protein was added to DEVD-CHO solvated in DMSO in 96-well V-bottom plates, sealed with tape, and incubated at room temperature. 22.5 μL and 90 μL aliquots were transferred to black-well plates in duplicate, and assayed with 50-fold molar excess of substrate. For kinetic measurements, 100 nM protein was assayed with the range of 0-500 μM DEVD-AMC over the course of seven minutes. Assays were performed at 37°C in 25 μL and 100 μL volumes in 96 or 384 well microplate format using a Molecular Devices Spectramax M5 spectrophotometer. Initial velocities versus substrate concentration were fit using Prism software (Graphpad Software) to determine kinetic parameters K_m and k_{cat}.

Enzyme activity recovery assays were carried out using WT and R210C/C246S mutant protein buffer exchanged to remove DTT into pH 7.5 10 mM sodium phosphate buffer using Viva Spin 3K concentrators (Sartorius Stedim Biotech,). Working at intracellular reduction conditions *in vitro* proved challenging, due to the propensity of caspase-7 active site to form mixed disulfides with glutathione. As a surrogate, we selected cystamine as the oxidant and dithiolthreitol or 2-mercaptoethanol as reductants. The reduction potentials of these four redox systems have been shown to be similar[112,113]. Protein was incubated in 96-well format in a range of cystamine concentrations (300 μM - 5 μM; 0 μM control) for a period of one hour. Half the volume was subsequently transferred to conditions with a final concentration of 10 mM dithiothreitol (DTT) and assayed for activity. DTT's redox chemistry is at equilibrium with cystamine at pH 7.9, supporting our use of 10 mM DTT as an adequate substitute for a highly reducing intracellular environment.

A similar ratio of oxidant to protein was used to treat caspase samples for caspase activity assays, immunoblot detection, mass spectrometry and crystallography. Caspases were then diluted to the concentration appropriate for the individual type of analysis.

Gel electrophoresis and immunoblot detection

WT and R210C/C246S protein was buffer exchanged using above method into pH 7.5 10 mM sodium phosphate buffer, then divided into two samples and incubated in the presence or absence of 300 μM cystamine for 2 hours. Samples for non-reducing gel electrophoresis were boiled for 10 minutes in Gel Loading Buffer (New England Biolabs) lacking DTT, and loaded onto 16% polyacrylamide gels. Immunoblots of the subsequently separated proteins were performed using antibodies specific for caspase-7 large subunit (SIGMA Monoclonal Anti-caspase-7 C1104) and small subunit (SIGMA Anti-caspase-7 PRS3465) and compared against known size standards (Kaleidoscope protein marker, BIO-RAD). False color merged photographs were generated by contrast enhancement then selective color replacement (Adobe Photoshop).

Protein mass spectroscopy

Protein for mass spectrometry was prepared at 18 μM in 120 mM NaCl, Tris pH 8.0 and 2 mM DTT with or without 1.5 mM cystamine. Mass spectra were acquired on a QStar-XL (MDS Sciex, Toronto, Canada) hybrid quadrupole/time-of-flight instrument. Reverse phase liquid chromatography (Agilent 1100 LC) using a C4 column was utilized to desalt and separate excess cystamine. Column output was infused directly into the ESI source at a rate of 0.5ml/min using N_2 as a nebulizing gas. Protein was eluted over a 10-minute gradient from 0%-100% acetonitrile and monitored using UV and TIC. A single protein peak eluted between 17-18 minutes at the end of the gradient. Data was analyzed using

Analyst QS (Applied Biosystems) with BioAnalyst extensions. All spectra deconvolutions were gated for proteins between 10 kD and 30 kD, with m/z ratios between 700 and 1700. The Mass spectrometers are housed in the Mass Spectrometry Center of the University of Massachusetts at Amherst.

Caspase-7 R210C/C246S crystallization and x-ray data collection

To prepare R210C/C246S crystals, protein in a buffer containing 120 mM NaCl and 20 mM Tris pH 8.0 was concentrated using Millipore Ultrafree 5K NMWL membrane concentrators (Millipore) to 6.3 mg/mL as assessed by absorbance at 280 nm. Cystamine was added to a final concentration of 1.25 mM. Crystal trays were setup at room temperature and grown in 3 μL hanging drop with mother liquor consisting of 2.1 M sodium formate, 300 mM and sodium citrate pH 5.0 in a 2:1 ratio of protein to mother liquor. Crystals grew to a maximum of 180 μm in 3 days at 20 degrees. Crystals were cryoprotected in 20% ethylene glycol in mother liquor with a 60 second incubation, then frozen by rapid immersion in liquid N_2. Data were collected using synchrotron radiation at APS 24-ID-C over 90 frames with 1° oscillations and showed diffraction data to 2.86Å. Indexing and integration were carried out with MAR XDS[114]. Data were Scaled using SCALA[74]. Data for the low radiation damage crystal were collected using with 1.5° degree oscillations. These data were indexed and integrated using HKL2000[73] and the data were scaled using SCALA[74].

Structure determination

Phase information was obtained by molecular replacement using 1F1J as the search model with residues corresponding to L2' omitted during searches with PHASER[115]. Removed residues were rebuilt into unambiguous density using COOT[116], then refined

using Refmac[117] restrained refinement. Water and formate molecules were inserted manually and checked for stereochemical appropriateness. The final refined model contains residues 57–196, 210–303 for chain A and 357–496, 510–605 for chain B. 103 water molecules are assigned chain O, and 7 formates assigned chain F.

Coordinates have been deposited in the PDB with accession code 3R5K

CHAPTER IV

COMPUTATIONAL PROTEIN RE-DESIGN USING EGAD

Abstract

Current tools employed in the study the role of individual caspases in apoptosis lack specificity, and thus a novel approach by an orthogonal effector will address such a limitation. In this work we propose to develop an orthogonal switch technology based on a computational redesign of the allosteric site at the dimer interface in caspase-7. This will allow for specific control over the function of caspase-7. Such a technology would facilitate the study of caspase-7 target substrates in cellular cultures, without necessitating the use of potentially misleading knock-out variants. The use of a computational system allows for the rapid sampling of millions of possible amino acid and molecule permutations on very short timescales. This high throughput sampling has allowed rapid *in silico* evolution of a small molecule binding site which was the objective of this part of the thesis. In the future, these designs can be tested by enzymatic studies and crystallographic methods.

Introduction

Design of orthogonal controls would address the lack of specific inhibition tools available for the study of various protein families. The ultimate goal of this approach is to develop a robust design and validate the technology on a whole range of caspases. This will allow us ultimately to spread this technology to other allosterically controlled proteins. The goal of this project was to use computational protein design methods[118] to redesign the

cavity at the interaction surface of two caspase-7 monomers and introduce specificity for a small molecule which would act as an allosteric inhibitor for that particular caspase.

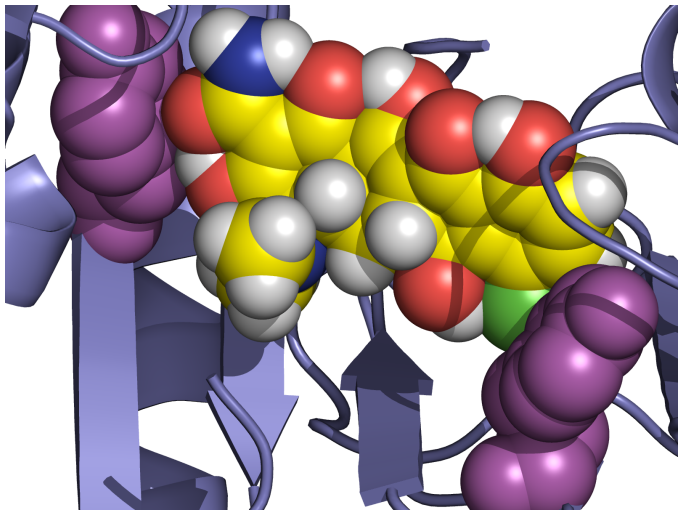


Figure 26: Small molecule effector placed ideally in allosteric cavity.

A small molecule (doxycycline) has been modeled into an ideal position in the allosteric cavity of caspase-7. The allosteric switch tyrosines (purple spheres) are sterically blocked by the small molecule from adopting the catalytically competent conformation.

The allosteric inhibition of caspase-7 by covalently tethered molecules has been studied[42,81]. The next logical step is to use a non-covalent reversible interaction to inactivate the protein. Using computational design as a low penalty, high throughput iterative design screen, we can combine established molecular biology methods, along with crystallography, to develop a model-to-protein-to-crystal design process of the protein. This will iteratively lead to a new protein variant that can be controlled orthogonally by means of a small molecule used in the design. Our computational design will begin with known crystallographic structures of caspase-7, which will be optimized for non-covalent binding of a tetracycline-class antibiotic molecule. In this optimization, the amino acid sequence that can best recognize the small molecule is computed. After the *in silico* design, the protein can be expressed in *E. coli* and purified for biochemical

characterization and crystallization. The resultant crystallographic data will allow refinement of computational models based on actual structural information, and perhaps elucidate new pathways for allosteric inhibition by small molecules.

Directed evolution is the best analogue for a computational design process. It must be said that there have been more successes with using directed evolution to engineer in new functions to proteins. However, as opposed to directed evolution, computational designs allow for screening of 10^{24} (13 sites per monomer, 17 potential amino acids at each position; $26^{17} \approx 10^{24}$) potential variants per simulation. To effectively screen similar numbers of variants using directed evolution is impossible with current technologies, which allow for screening of 10^{9}-10^{13} variants maximally. Even to screen these smaller numbers of variants would require the development of a robust reporter of caspase inactivation, combined with rapid screening methodologies. However, the use of such technologies would require significant additional capital and time investment as compared to a computational trial. Thus computational protein re-design is the most cost effective rapid initial screening method for the introduction of new binding motifs. Furthermore, the prospect of combining directed evolution with computational design can increase the likelihood of success dramatically.

Success in protein design and redesign efforts lies in the availability and openness of the design platform. Many efforts have been duplicated in the field of computer-based protein design, however, the growing popularity of the open source software movement may stem this tide. Programs like DEZYMER[119] have been historically difficult to obtain, with source code not being made available. In the absence of the original source code, the inner workings of the energy functions and parameterization are hidden from the

operators, keeping the software more single purpose. However, the availability of source code does not automatically guarantee a better working environment. Rosetta[56], by far the most popular current protein-design package, does offer source code, however the documentation of the source code is reportedly minimal, leaving the user to reinvent the wheel in the process of customizing the software for a novel purpose. On the other hand, EGAD[118], while not in active development, has been released freely, with extensive documentation to allow for easy customization by novice computational chemists. In addition, at the time this project was implemented, EGAD was one of the few programs that could design proteins against a small-molecule ligand. It is therefore important to not only consider the success of the tool used for protein design, but also the availability of documentation.

Existing EGAD toolkit

EGAD, or **E**GAD!, a **G**enetic **A**lgorithm for Protein **D**esign, was developed initially Navin Pokala under the direction of Tracey Handel[118], and has since been updated to EGAD Library, by Arnab Chowdry[120]. EGAD Library is based on EGAD's force-field implementation and according to the application's literature, "is intended for end users with programming experience, written in C++, for the development of new protein design applications." As an open source and well-documented application, EGAD allows novice programmers to develop and extend the current toolset for the users particular interest. This full access to the development of the program is very advantageous over other more black-box suites, as features and bugs can be worked out. The use of this method allows for a high throughput design from the computational model to the expressed protein, and ultimately the solved crystal structure of the allosterically inhibited caspase.

EGAD's computational approach relies on structural protein data in the form of a set of coordinates in a PDB file along with the structural data of the small molecule. The initial starting location of the small molecule, henceforth referred to as the ligand, in relation to the protein is preset by the user. Molecular visualization programs such as PyMol[79] or COOT[116] can be used to design this initial small molecule placement based on existing knowledge of the allosteric switch in caspase-7[42]. The selection of residues comprising the potential binding cavity is similarly chosen. A variable residue can either have full mutability, in that EGAD can choose the best amino acid identity, save for Gly, Pro, Cys, and rotamer for that particular position. The second option is to maintain amino acid identity at a position, while allowing that particular position to select from the rotamer library to optimize the position.

The selection of ligand is restrained by two criteria. First, desired ligand should have no appreciable effects on wild-type caspase-7, and secondly, the ligand must be amenable to computational design in that the molecule has few rotatable bonds and the molecule must be of appropriate size for the cavity. One potential family of allosteric effectors is the tetracycline antibiotic family. Preliminary caspase-7 inhibition studies were carried out in the Hardy lab to determine possible tetracycline derivatives that are appropriate for use in design. Besides tetracycline, doxycycline, demeclocycline and minocycline were tested for their inhibitory effect on the wild type caspase-7 protein.

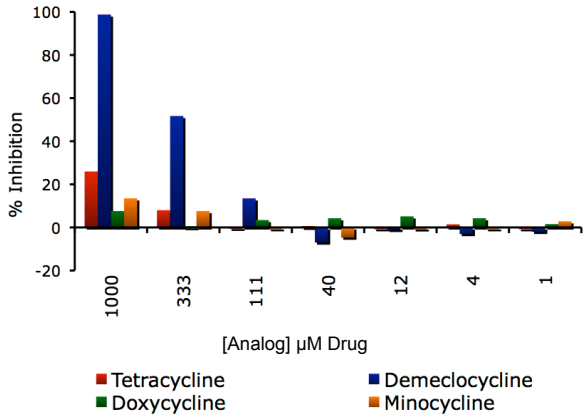


Figure 27: Effects of tetracycline analogs on caspase-7 activity.

Wild-type caspase-7 was incubated with increasing concentrations of the indicated inhibitors for one hour. The activity of caspase-7 was then tested in a kinetic cleavage assay using the fluorometric substrate DEVD-AMC. The amount of inhibition is relative to a DMSO control.

(Figure 27). Demeclocycline showed inhibition of the protein at relatively low concentrations through an undetermined mechanism, and so was ruled from further investigations. Doxycycline and minocycline both had similar inhibition curves to that of tetracycline and thus were considered for further study using EGAD. Doxycycline has been chosen as the most likely candidate for allosteric effector.

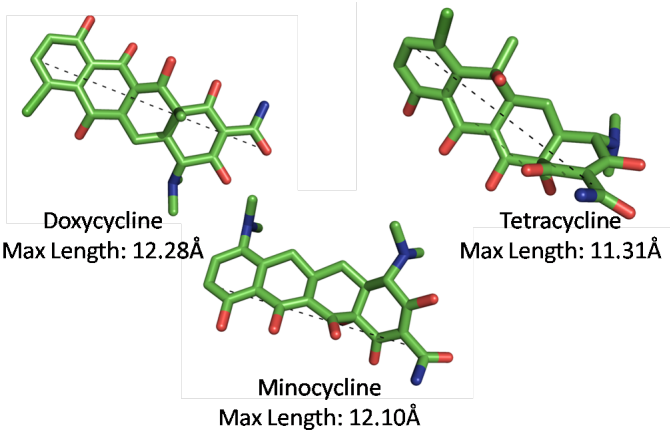


Figure 28: Potential small molecule effectors

The length of the potential small molecule effectors is of interest since the effector must effectively fill the cavity of caspase-7.

To produce a small-molecule starting model for future designs, the low energy structure of each small molecule was determined computationally using Hartree-Fock type *ab initio* calculations to the RHF-6- 31+g basis set using Gaussian 03[121]. The reliability of this calculation was determined by the calculation of the minimum energy structure of tetracycline, and comparing it to that of a solved crystal structure that included tetracycline in complex with the ribosome, with the two results being very similar.

EGAD Computation Setup

EGAD is a command line driven program with no graphical user interface, and so all commands and input must be hand coded into input files. The input file generation for EGAD is by far the most time consuming of steps prior to calculations. The small molecule's molecular description consists of a ligand input file and a ligand parameter file, which contains all the atomic coordinates, the atom types, atom charges, and inter-atomic connectivity. Atomic coordinates are represented as X, Y, Z and are in the same space as the protein. Atom type is for the OPLS force-field[122] and the atom type is assigned manually at the discretion of the user. The atom type depends on the atomic environment in which the atom resides. Atomic connectivity is self-explanatory and charges are calculated during the energy minimization step of the geometric optimization in the *ab initio* calculation of the structure of the ligand. Rotatable bonds can be described in EGAD, however this feature is of limited use as the bonds can only be described as a set of three rotations about a fixed atom.

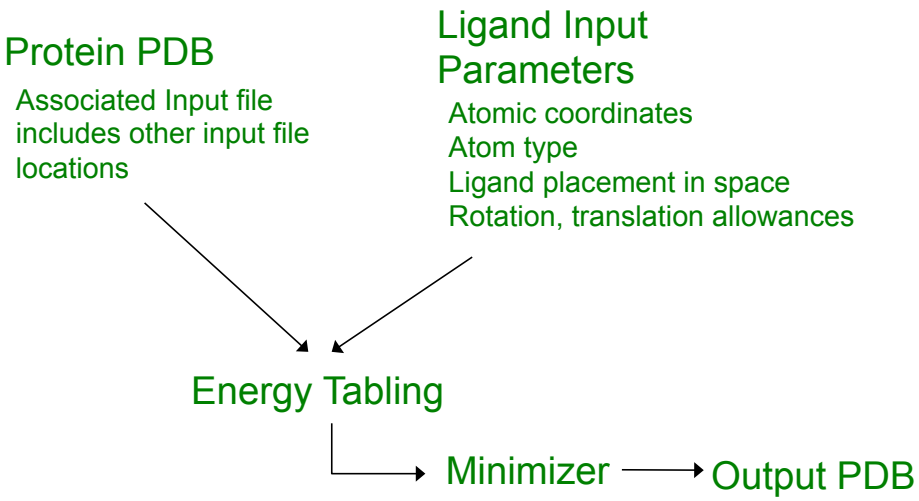


Figure 29: EGAD operation flow

Files containing atomic coordinates and atomic descriptors for both the protein and small molecule ligand are input into EGAD for energy tabling. The energy table is then loaded into an energy minimizer, which uses DEE and Monte Carlo methods to search for an energy minimum. Once sufficient solutions are found, the data is then output as coordinate files in the form of PDBs.

The second file that describes the ligand is the ligand input file. This file calls the ligand parameter file at the start of the calculation, and includes the coordinates of the centroid of the molecule. The centroid is calculated by finding the mean of all the discrete coordinates of the atoms within the ligand. The input file includes the variables for allowed ligand translation (spread_radius) and allowed ligand rotation (XYZ_Range). The allowed ligand translation variable sets how far from the initial centroid coordinates the calculation is allowed to move the molecule (in Angstroms), while the ligand rotation variable allows for control of the rotation range (in radians) in the calculation.

EGAD General Energy Function

During computational redesign, each variant is scored by the EGAD energy function. EGAD's energy function is a linear combination of the calculated energy of the statically held backbone atoms (ΔG_{i_bkbn}), the energy of the particular rotamer at position i ($\Delta G_{i\_internal}$) and the pairwise interaction of rotamers at positions i and j ($\sum_{j>i} \Delta G_{ij}$).

Each portion (ie each term) of the equation is further expanded to the vacuum energy and solvation energy components. The vacuum energy consists of the VdW, Coulombic, Torsional correction terms, while the solvation energy consists of Generalized Born calculation.

(Note: ΔG is the modified Gibbs free energy equation, while the E represents a free energy approximation, as the entropy term has been omitted in the calculation.)

$$\Delta G = \sum_i \left(\Delta G_{i\_\text{internal}} + \Delta G_{i\_\text{bkbn}} \right) + \sum_i \sum_{j>i} \Delta G_{ij}$$

Figure 30:Total Energy calculation term

$$\begin{aligned} E_{forcefield} &= E_{vdw} + E_{coulombic} + E_{torsion} \\ &= \sum_{i=1} \sum_{\substack{j>i \\ i,j>3 \text{ bonds apart}}} \left[4\omega_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) \right] \\ &\quad + k_c \sum_i \sum_{\substack{j>i \\ i,j>3 \text{ bonds apart}}} \left[\frac{q_i q_j}{\varepsilon_p r_{ij}} \right] \\ &\quad + \sum_{i=1} \sum_{\substack{j>i \\ i,j=3 \text{ bonds apart}}} \left[\begin{array}{l} 2\omega_{ij} \left(\frac{\sigma_{ij}^{12}}{r_{ij}^{12}} - \frac{\sigma_{ij}^{6}}{r_{ij}^{6}} \right) + \frac{1}{2} k_c \frac{q_i q_j}{r_{ij}} \\ + \sum_{n=1}^{3} V_{n_{ij}} \left(1 + (-1)^{n+1} \cos\left(n\phi_{ij} \right) \right) \end{array} \right] \end{aligned}$$

Figure 31: van der Waals, Coulombic electrostatic potentials and torsional potential

It is worthy to note that bond stretching and angle bending terms are not taken into account in this calculation, because the model assumes a rigid backbone and uses idealized rotamer conformations. To reduce computational overhead, long range interactions (greater than 10Å) are ignored. Dielectric constants were chosen empirically

to a value of 8.0. Again, we have atoms *i* and *j* , with r_{ij}, with charge q_i, being the radial distance between the two, and φ_{ij} , σ_{ij}, ω_{ij} being their dihedral angle, VdW equilibrium distance and well depth respectively. Lastly, in the torsional potential term there is a $V_{n_{ij}}$ term which represents the Fourier series coefficients for torsional energies.

$$\begin{aligned}\Delta G_{electrostatics} &= \Delta G_{Coulomb} + \Delta G_{Born} \\ &= \Delta G_{Coulomb} + \Delta G_{self} + \Delta G_{pair} \\ &= k_c \sum_i \sum_{j>i} \frac{q_i q_j}{\varepsilon_p r} + k_c \frac{1}{2} \tau \sum_i \frac{q_i^2}{\alpha_i} + k_c \tau \sum_i \sum_{j>i} \frac{q_i q_j}{f_{Born}}\end{aligned}$$

where:

$$\tau = \frac{1}{\varepsilon_w} - \frac{1}{\varepsilon_p}$$

$$f_{Born} = \sqrt{r^2 + \alpha_i \alpha_j}$$

Figure 32:Generalized Born model used in EGAD

In the equation above, the Born radii are calculated from the supplied protein structure. Where k_c is Coulomb's constant, *i* and *j* are the atoms, with *r* being the distance between said atoms. We take into account two dielectric constants, that of the molecule (ε_p), and that of the surrounding solvent (ε_w), along with the individual charge of each atom *i* (q_i) with a Born radius of α_i. The screening interactions are described by f_{GB}.

The energy terms only apply to the protein, and not to the small molecule, the small molecule is a rigid structure, whose conformation was determined by energy minimization using *in vacuo ab initio* methods, specifically the 6-31+g basis set of the restricted Hartree-Fock calculation by Gaussian 03[121].

EGAD Efficiency

EGAD's efficiency stems from the architecture of the calculation. The energies of all the possible combinations of rotamers and mutants for the selected amino acids are calculated and stored. The precalculation of energies, though space inefficient, allows for rapid energy lookup.

The low energy search algorithm employed by EGAD in our experiments uses a stochastic approach found in Monte Carlo simulations. The Monte Carlo search allows us to sample rapidly a large energy landscape that has been parameterized in the previous tabling step. Monte Carlo methods sample the local energy space until a local minimum is found, and with large number of attempts, this local minimum is assumed to be the global energy minimum. Dead end elimination has also been implemented in the energy minimizer, though its use is limited to initial pruning prior to Monte Carlo minimization. The greatest limitation is the necessity to maintain rigidity of the backbone atoms during the calculation. Because of the complexity of including backbone flexibility in the calculation, certain restrictions and compromises must be made to make the calculation feasible. The van der Waals radius scaling factor, is variable between 90% and 95%, to account for small backbone rearrangements.

SURFNET

The torrent of data following an EGAD run is difficult to parse without context. In a typical calculation, between five and 100 output structures can be generated, each with a different final energy and amino acid composition at the variable sites. The minor energy variations between each solution are not necessarily meaningful, as composite

assumptions within the calculation lead to large error bars. Thus a secondary screening methodology is required to further determine the best result.

To create the best binding cavity for the small molecule, additional space surround it must be minimized. As a secondary screen, the output from EGAD is processed via SURFNET[123] which generates molecular surfaces of molecules, and can find voids between the protein and the small molecule. SURFNET is freely available online (http://www.ebi.ac.uk/thornton-srv/software/SURFNET/) to academic institutions, and has been installed on the lab linux workstation Solitude. SURFNET processes a single output file individually, as determined by the input parameter file associated with SURFNET (surfnet.par), and outputs a series of files that contain geometric information of the voids between the protein and ligand, among others. SURFNET has other functions such as cleft detection, and numerous geometric rendering options, however these are beyond the scope of this thesis.

Running EGAD & SURFNET

EGAD runs are composed of a series of terminal commands, and can be easily batched together. For data separation simplicity, each run of EGAD performed is within a serialized directory structure. For instance, the directory *doxdes3.2.11* indicates a design run using the doxycycline ligand, third generation of major design principals, second set of experiments, 11th variant on input parameters. This directory structure is meant as a guide for simplification of data analysis, and specifics are likely to change from run to run. Each directory (and therefore individual run) requires the following files.

common_inputs directory

This directory contains the rotamer libraries, Van der Waals residue parameters, torsional parameters and atom specific information as part of the Van der Waals radii for the OPLS forcefield, and test points for the SASA calculation. As this directory may be modified between runs, it is useful to reference a local copy within the directory structure for each run. These files are called by the protein input file component of the EGAD setup.

etable.cmdline

Etable.cmdline contains the flags for EGAD that specify the protein input files, the ligand input file, and whether WT residues are allowed.

```
allow-wt
egad-input      1SHJ_A.input,1SHJ_B.input
ligand-inputs   dox.liginput
```

Figure 33: etable.cmdline

In this example, the EGAD protein input files are called by 1SHJ_A&B.input, the ligand file is called by dox.liginput, and WT residues and rotamers are implicitly allowed.

1SHJ_A.input

For this example, we are using separate protein files for each side of the dimer to allow each side to be addressed and varied independently. The first file called contains calls to files and parameters for the protein component of the run. The critical flag of the file, *TEMPLATE_PDB* refers to the location of the PDB input file. This file should be in the directory of each run, and in this example, specifies the A chain of 1SHJ.pdb

```
#### Variable position definitions ####
# This section defines which positions will be allowed to mutate
# during the design calculation. Its a simple table of position indices
# (from the PDB file) and the residues allowed at that position.
#
# Any positions not mentioned here are assumed to be "fixed", that is
# to say they will not move during a design. Any positions defined to
# be the same as their wild-type counterpart will be allowed to change
# rotamers.
#
# The special keyword "all" will allow any residue type (except GLY,
# CYS, and PRO).
VARIABLE_POSITIONS
    223         TYR  # ONLY ROTOMER CHANGE
END

#### Energy function settings ####
# Each of these flags sets whether a portion of the energy function
# should be used in the design calculation. A '1' means that it will
# be included, while a '0' means it will not.
VDW_FLAG          1
COULOMB_FLAG      1
TORSION_FLAG      1
SASA_FLAG         1
GB_FLAG           1
HBOND_FLAG        0
REFERENCE_FLAG    1

#### Energy function parameters ####
# Various parameters that modify how the energy functions behave.
# At this point, I just use this to define what scaling factor to
# apply to the VDW radius.
VDW_RADII_SCALE 0.90

END
```

Figure 34: Protein component input parameters

The first file loaded by etable.cmdline contains protein component input parameters such as the location of the PDB that stores the geometric coordinates (*TEMPLATE_PDB*), and the variable positions within that PDB.

The *VARIABLE_POSITIONS* flag refers to the residues that are specified for computation. In the example (Figure 34), only residue at position 223 is included, forced into tyrosine (TYR), and only with rotameric freedom. This calculation will allow the ligand to move, position 223 to sample rotamers, with the remaining residues held fixed. If a specific mutation is called for, or only specific residues allowed, the allowed residues

can be listed in comma delimited format. The flag 'all' will allow any residue type, except glycine, cysteine and proline (Figure 35).

```
VARIABLE_POSITIONS
    147         all  # Glu
    148         all  # ASN
    159        # ILE
    160     PHE,TRP,TYR # ASP
    213         ILE  # ILE
    214         all  # Pro
    215         all  # Val
    216         all  # Glu
    217         all  # Ala
    221        PHE,TRP,TYR # PHE
    223         TYR  # ONLY ROTAMER CHANGE
    290         all  # CYS
    292         all  # VAL
    294         all  # Met
```

Figure 35: Variable positions in EGAD

Variable positions explicitly called by the protein parameter input file. Positions not listed are modeled as listed by the PDB geometric coordinates, but not varied, nor allowed to move.

Notations after hash (#) are ignored by EGAD, and can be used as comments for the user. The 1SHJ_B.input file only contains the location of the B PDB, and variable positions in that PDB.

Machinefile

This file is only called when running in an MPI parallel processing cluster environment[124], and contains machine names on which to run. All EGAD simulations to date have been run on either the generously provided Handel *pandemonium* cluster at UCSD or the compbio cluster at the University of Massachusetts, Amherst.

Dox.liginput

The contents of this file refer to the ligand parameter file, and also mark the initialization coordinates for the ligand. Under the location flag, the coordinates listed are the center of the ligand, and can be calculated by averaging the X Y and Z components separately

from the ligand parameter file. The Spread Radius refers to the distance in Angstroms the ligand moves from the initial location, while the grain is the number of steps the ligand will make in each axis. The rotation and range flags indicate the number of rotations to make about each axis, and the range, in radians, which rotations should be made.

Dox.ligparm

The ligand parameterization is contained within this file. In this specific example, the file contains necessary information to define doxycycline. Layout of this file is reminiscent to a PDB file, in that the file is in a space-delimited format. Column 1 contains the flag for coordinates, followed by atom number, atom identity, then atom type. The atom type is for the OPLS forcefield and needs to be determined by finding the closest atom type in the forcefield file that describes that atoms chemical situation. The column composed primarily of 0's refers to the rigidness of that atom relative to other atoms connected to it. Following this column are three columns of X Y and Z coordinates that allow atom distance information to be calculated. The final column is the charge associated with the atom, and this information is taken from an *ab initio* energy minimization calculation. It is important to note that to move the ligand to a new staring position requires that new coordinates be entered into this file, and a new center of coordinates be calculated.

tabling.sc

This script file builds the master tabling input file from components outline above and starts the program called MPIPolymerTable, with the output file being generated in the root directory of the script starting location.

```
#!/bin/bash
# Tabling Script, by Witold

echo "Running Tabling"
mpirun -all-cpus -machinefile machines ~/EGAD_Clients/bin/MPIPolymerTable --CLPL
oad=etable.cmdline &> etable.log &

echo "The script should be running, I'm printing the log file as it is written."
echo "push CTRL+C to escape"

tail -f etable.log
```

Figure 36: tabling.sc start script

Tabling initiation script that ensures the proper syntax of a MPI distributed computing run to be executed using the etable.cmdline and machinefile files as inputs, while generating a log file (etable.log) in the same directory.

Stopping this script with the escape character during tabling will not stop the tabling run, only halt printing of the tabling log file. To stop tabling, the job must be killed by PMID in the linux process manager on each machine (node) running the tabling. Running two simultaneous tabling jobs in the same directory can lead to corrupted output files, and is generally bad practice. Depending on the input parameters, job size, and available computing power, tabling can take between 6-12 hours. Though the whole set is ostensibly parallelizable, the single most computationally intensive segment is the ligand parameterization, which cannot be spread between machines. It is therefore recommended that tabling be undertaken on one machine. Output files generated end in the file extensions .etable and .groups.

Monte_carlo.cmdline

Once tabling is completed, search for the minimum energy value can be done using the MonteCarlo program of EGAD. The random seed is an important variable, and must be changed between each run of the same input to ensure MonteCarlo produces independent results (Figure 37). The number of runs, saved and number of steps is self explanatory, as these parameters determine the length of the random walk, and the number of top

solutions to be saved. The penultimate flags in this file refer to the output files from tabling. These files should reside in the same directory as the MonteCarlo run, and so file paths are unnecessary. The final flag in the input file refers to the number of DeadEndElimination steps to take. This flag is optional.

```
random-seed     116200268
num-runs        100000
num-saved       1000
num-steps       10000
etable  dox_caspase.etable
groups  dox_caspase.groups
eliminate 2
```

Figure 37: monte_carlo.cmdline

The monte_carlo.cmdline file houses pointers to the tabling output files (etable and groups) and specifies the number of runs, how many of those runs are saved, and how many steps per run are to be in each Monte_Carlo run. *Eliminate 2* refers to the rounds of DEE prior to the start of Monte_Carlo.

montecarlo.sc

The montecarlo.sc script is very similar to tableing.sc (Figure 36), in that it invokes the MonteCarlo program with the proper input files listed (Figure 38).

```
#!/bin/bash
# Monte Carlo Minimizer Script, by Witold

echo "Running Minimizer"
echo "Monte Carlo Simulation"
echo monte_carlo.cmdline
mpirun -all-cpus -machinefile machines ~/EGAD_Clients/bin/MPIMonteCarlo --CLPLoa
d=monte_carlo.cmdline &> monte_carlo.log &

echo "The script should be running, checking monte_carlo.log for results"
echo "Push CTRL+C to quit log file"
tail -f monte_carlo.log
```

Figure 38: montecarlo.sc start script

Similar to the tabling script, this script initiates the minimizer step of EGAD while in a MPI controlled multi node environment.

Struct.sc

The output file from the MonteCarlo program (monte_carlo.log) contains references to the atomic positions that represent the minimum for the run. The script

print_structuresnue.py is a python script whose input syntax takes the monte_carlo.log file, the protein file and the ligand file to generate PDBs of final output structures. Struct.sc calls print_structuresnue.py two times, to generate SASA volumes for the protein with and without the ligand in the final conformation present. This feature has been primarily deprecated due to the more accurate SURFNET being employed, though the additional information from a SASA water roll is useful in some instances where the ligand is protruding from the allosteric pocket.

PRESURF.sc

The SURFNET program is run after PDB model generation. Modification of the surfnet.par file is required to specify the number of atoms within the protein and the ligand. SURFNET then calculates the volume between the two bodies using spheres between 1.0Å and 4.0Å. The additional options afforded by SURFNET are beyond the scope of this thesis, and can be best referenced at the SURFNET website. The presurf script generates a new directory, *surfout*, where the outputs from the SURFNET run are subsequently written. Presurf.sc removes all previous run outputs generated by SURFNET prior to starting a new SURFNET calculation.

A successful SURFNET calculation will yield many output files, but the critical files for EGAD ligand/protein spacing analysis are surfnet.pdb and volumes.pdb. Simple visualization of the remaining space surrounding the ligand can be accomplished by loading the output PDB from EGAD into a molecular viewer such as PyMol along with surfnet.pdb. This will generate two sets of MAP points, one corresponding to the small molecule, the second corresponding to the surrounding space.

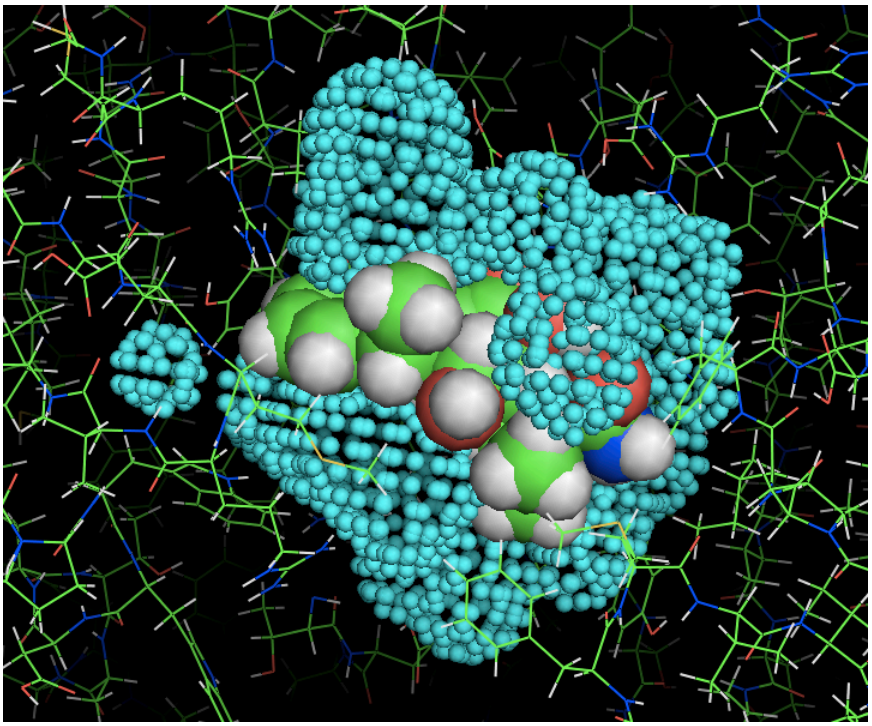


Figure 39: SURFNET output displayed in PyMol

An output solution from EGAD and SURFNET rendered in PyMol. The extra space (cyan spheres) surrounds the ligand (green spheres). Hydrogens not seen in the crystal structure have been added where chemically prudent.

To quantify the volume remaining in the cavity, a simple calculation must be done with figures from volumes.pdb. During the SURFNET calculation, SURFNET calculates the total volume of the ligand+protein, protein only and ligand only. Remaining volume of the cavity is calculated by subtracting the ligand volume from the total volume of the complex, then subtracting the result from the volume of the protein only.

Initial EGAD Designs using Tetracycline

Initial EGAD designs were undertaken in a close collaboration with Dr. Arnab Chowdry and Dr. Tracy Handel of UCSD. Tetracycline was chosen as a ligand design target at that time, as an energy-minimized structure for doxycycline was not yet available.

Residues selected for inclusion in the first round of design were chosen for their proximity and potential interaction with the tetracycline (Figure 40 A, Table 6).

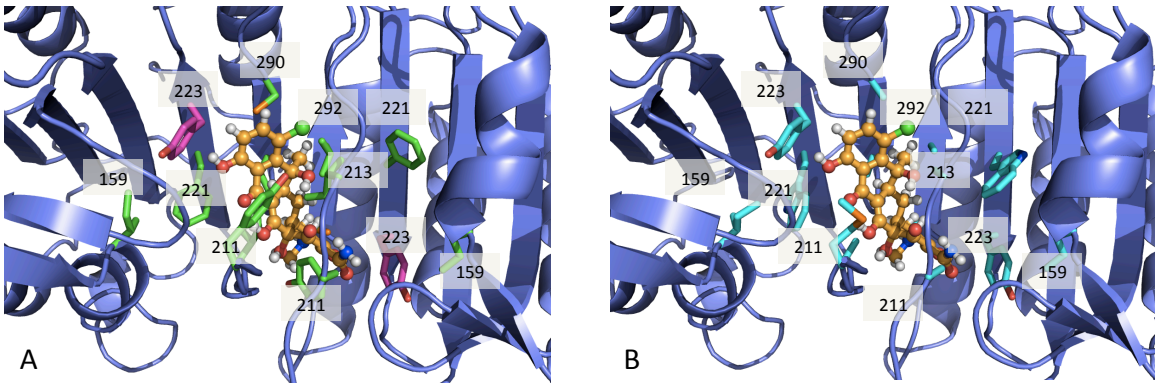


Figure 40: Des1.0 binding ligand interaction pocket

(A) The native interaction pocket with residues included in design (green) labeled. Tetracycline (orange) is in the conformation predicted by EGAD. (B) Des1.0, the designed output from EGAD.

Residue #	WT	EGAD Result		Expressed Variant
		A	B	
159	Ile	Ile	Ile	Ile
211	Tyr	Met	Ala	Met
213	Ile	Ala	Ala	Ala
221	Phe	Trp	Trp	Trp
290	Cys	Ala	Ala	Ala
292	Val	Ala	Ala	Ala

Table 6: Design results for Des1.0 at mutated positions

As mentioned in chapter 1, the free rotomeric freedom of the tyrosine at 223 is critical for activity. As such, the results from the first design were troubling in that the adjacent residue to the allosteric switch tyrosine 223 was a tryptophan. The bulkiness of the resultant residue, while beneficial to the overall design, posed a potential problem if the tryptophan restricted the movement of tyrosine 223. As a first step, the single point

mutant F221W was generated (Figure 41 A) and its activity measured relative to wild type enzyme (Figure 41 B). No appreciable difference was noted in both expression of the F221W variant, nor in the activity.

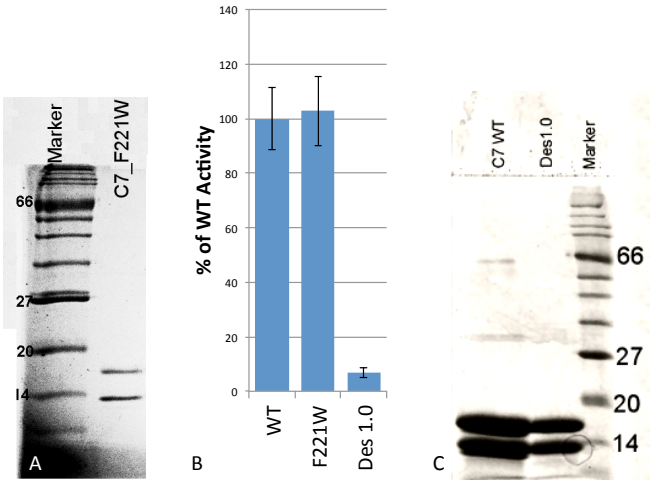


Figure 41: Expression and validation of F221W & Des1.0

Expression and activity validation results for the intermediate design mutant (F221W) and the full first generation design (Des1.0). The expression of F221W yielded a protein with expected molecular weights (A), and activity recorded was comparable to WT (B). The full mutant design, (Des1.0) showed significantly decreased activity relative to WT (B), but behaved just as WT in a gel assay.

The positive result with F221W led to the generation and expression of the full design mutant. However, the EGAD computation resulted in an asymmetric design, which due to limitations of current caspase-7 expression systems, was not available. Caspase-7, being a obligate homodimer would require the generation of two independently mutateable monomers. The generation of such a expression system construct was undertaken, the project remains unfinished. This technical problem required the generation of a design that would be symmetrical about the allosteric cavity. Conveniently, the EGAD solution was nearly symmetrical on both sides of the dimer, with only position 211 varying. Upon examination of the computational result, it was decided that a methionine would be beneficial to include on both sides of the dimer, as

opposed to mutation to an alanine. The position of residue 211 near the N-terminus of the L2' loop is considered mobile, as the L2' loop necessarily requires rearrangement during the substrate binding groove formation (Figure 6).

The full design (Des1.0) was expressed, purified (Figure 41), and assayed against WT to assess relative activity. Though the expression of Des1.0 resulted in a protein that behaved like WT in both purification and gel assay, the catalytic activity was severely impaired (15-fold reduced) relative to WT. However since Des1.0 was still a somewhat functional caspase-7, and given that the primary goal of the design is to engineer in sensitivity and binding for tetracycline, we tested this. To assay tetracycline-mediated inhibition, caspase-7 and Des1.0 were subjected to tetracycline to determine IC_{50} irrespective of the intrinsic activity (Figure 42). Des1.0 showed a 50% lower IC_{50} relative to WT, demonstrating a partial success in the computational design.

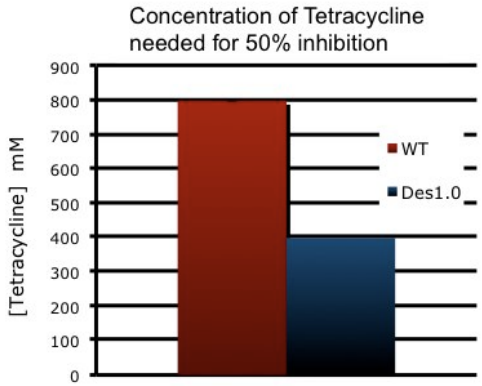


Figure 42: Caspase-7 vs Des1.0 IC_{50} to tetracycline

The sensitivity of Des1.0 relative to WT caspase-7 to the target tetracycline molecule is 50% higher, suggesting a successful first round of design using EGAD.

The severely lowered caspase-7 activity of Des1.0 lead to the investigation of L2' as a candidate for design by alanine scanning mutagenesis (Chapter 2). It was subsequently discovered that position I213 is of critical importance to the formation of the substrate

binding groove, and thus both positions 213 and 212 were removed from subsequent design initiatives.

Designs post Des1.0

The lessons learned during Des1.0 and further investigations to L2' loop's amenability for mutation guided later computational designs. Because doxycycline was 1Å longer than tetracycline, and could therefore reach between the two Y223 residues on opposite sides of the allosteric pocket, doxycycline was used in all subsequent rounds of design. The doxycycline structure was subjected to *ab initio* energy minimization in Gaussian 03[121] and geometric and chemical parameters were input to the dox.ligparam file. Additionally, residues 212 and 213 were included in the calculation, but were only allowed rotameric freedom.

A number of variables that included Van der Waals radii scaling, random seed changes, included residues, ligand starting position and movement were varied in numerous runs with a total of 90 successful solutions compared. A direct comparison of similar runs, as defined by the same residues being varied, can be made and trends from these results can be elucidated. In set one, with 42 solutions across 14 different setups we find where certain residues are preferred over others. For instance residue number 215 in chain A is preferred and 89% of the time, while other residues are selected here the remaining 11% of the time. However the preference of one residue over another is not always so overwhelming, as residue 292 is split between methionine, alanine, and arginine in almost equal amounts.

The strength of the asymmetric nature of the design is evident when one compares chain A to chain B. Using the example of residue 215, chain A differs from chain B in that chain A, glutamate is chosen while in chain B a tryptophan residue is selected 84% of the time. It is important to note that these consensus residue selections do not always translate across different input parameters.

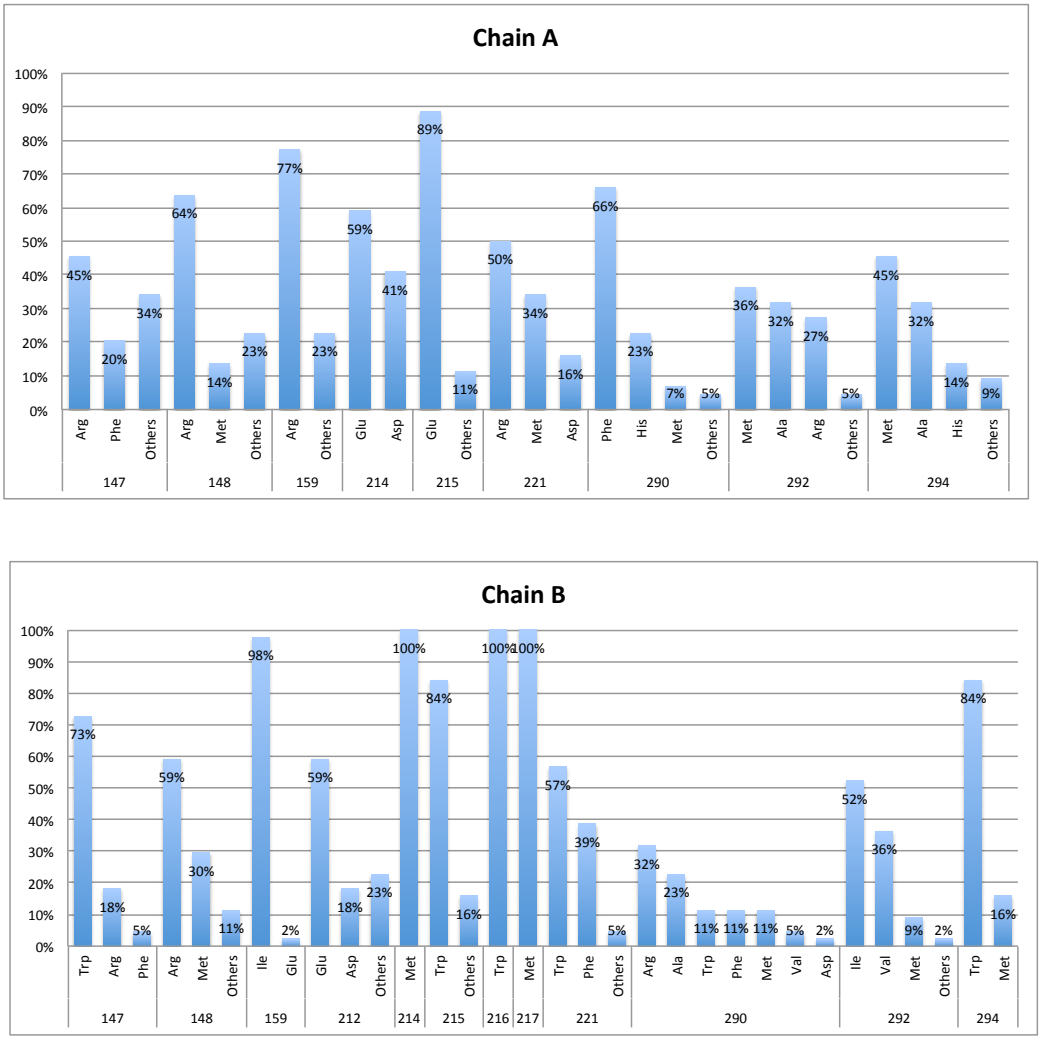


Figure 43: Consensus residues across one set of solutions. (N=42)

Percent selection of a specific residue at each the variable design position for EGAD design series 3.2.x. Percentages are score based on number of occurrences of that residue in all solutions from multiple runs.

A second set of results (48 solutions from 11 separate runs) differs from set 1 in that different residues were selected for mutation and the starting position of the small molecule was also changed. The reproducibility of EGAD as a design tool becomes evident when one compares two starting positions of the small molecule and the resulting consensus residues predicted by the computation. For instance, at residue 215 in chain A, the predicted consensus residue is a glutamate, while in the second set at the same residue position an aspartate is called for. This suggests that an acidic residue in this position is likely to be beneficial for binding of the small molecule. Trends such as these can be funneled into subsequent designs, such that position 215 is only allowed to vary to acidic residues (Figure 35). Finally, after a few rounds to select for not only residue chemistry, but also possibly residue identity, scoring based on residual space in the small molecule cavity can be undertaken.

No computational design result can be discussed without mentioning final energies calculated. During the minimization step, when a Monte Carlo simulation is searching for the lowest set of pairwise energies, EGAD uses a force field calculation in hartrees to score outputs. However, due to the variances and inherent approximations within such a calculation, comparison of the direct energy values between runs is not recommended. Generally speaking, output energies are on the order of -1900 to -2300 depending on factors such as Wan der Waal radii and final position of the small molecule are typical from an EGAD calculation. Higher energies, especially if positive, suggest steric clashing, which we could observe post-run.

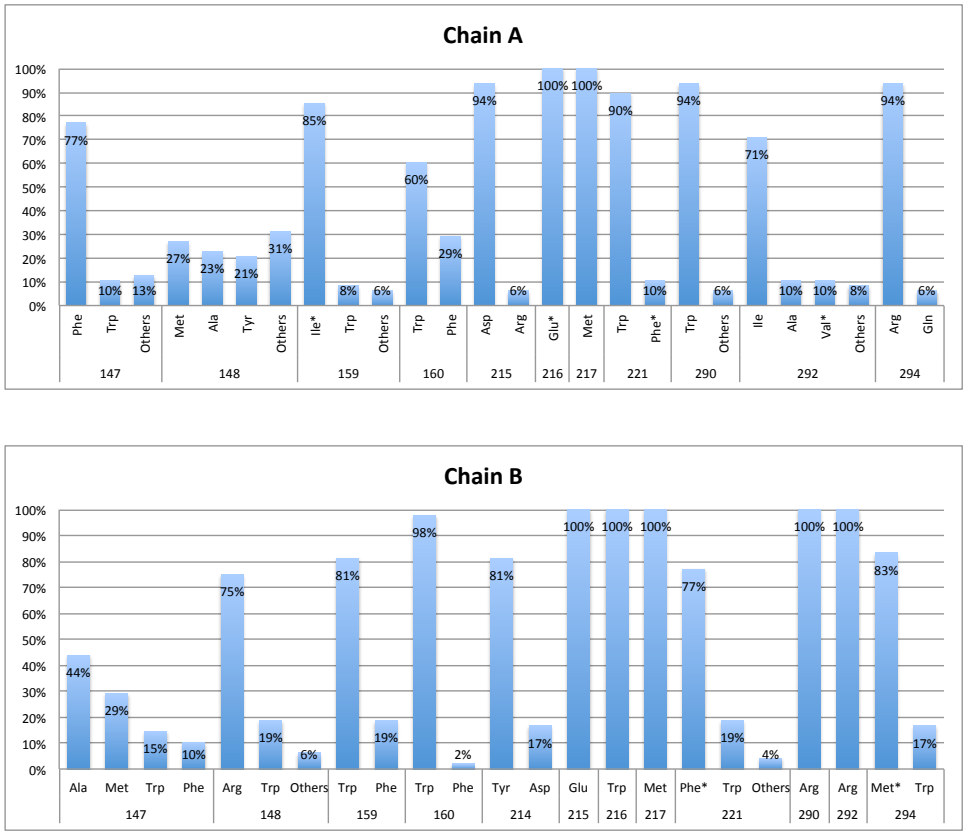


Figure 44: Consensus residues across one set of solutions. (N=48)

Second set of solutions (3.3.x series) based on a greater number of positions varied and a new starting position for the small molecule.

The remaining space surrounding the small molecule in each design varies based on the final ligand position, and the surrounding residues predicted. Selection of protein designs to carry forward for expression in *E. coli* must be made not simply based on consensus residues at specific positions but also their contribution to a smaller binding pocket. In the case that a residue does not fall into a consensus, the secondary screening for unfilled cavity volume using SURFNET will allow further refinement and selection of residues for mutation.

Using tools like those outlined above can greatly simplify the selection of mutants to test experimentally. Experimental validation of these designs has not been attempted due to

the inability to express and purify a heterodimeric version of caspase-7. Development of a heterodimeric caspase-7 has been undertaken by another graduate student in the lab.

The results of the calculations presented here, along with source code for EGAD and the scripts which automate the computation can be found in the /home/shared/EGAD directory on the Hardy linux workstation, solitude. As of the writing of this thesis, EGAD has fallen into disrepair and is not currently accessible on either the UCSD cluster or the compbio cluster. Installation and compilation of EGAD from source is possible, though directions for this task are beyond the scope of this thesis. SURFNET has been installed on the Hardy solitude linux workstation, and can be invoked using the command SURFNET in the directory in which the proper input files are located.

CHAPTER V

CONCLUSION

Caspase-7 is an excellent system to study, as the wealth of knowledge that has been gleaned in this work from the investigation of its loop movement, active site conformation, and allosteric control can be generalized as a model system for the study of other caspases, and even proteases in general. The outwardly unique characteristics found in caspase-7 have analogs within the protease family, and work on other proteins can be improved based on the findings presented in this thesis.

Loop or structural rearrangement upon binding of substrates or co-factors is not unique to caspases. The classical example is hemoglobin, where a tetramer undergoes a global conformation rearrangement as a function of binding oxygen[125]. In this example, there is a large conformational rearrangement change upon binding oxygen that has become a hallmark of the protein. Other systems have large conformational changes in their life cycle, such as HIV protease, whose flaps open and close depending on binding state[126], or influenza hemagglutinin, which seems to adopt a 'spring loaded' injection mechanism during its function[127]. All these structural rearrangements are required for the protein to function. In caspases, there is an intersubunit cleavage that is required for activation, implying that either the conformation of the zymogen is not sufficiently flexible to undergo catalysis, or that a structural rearrangement is necessary. The best metaphor for an executioner caspase is an unloaded pistol. Evolution has designed multiple fail-safes to ensure that caspases do not inadvertently activate. In our metaphor, the zymogen caspase is an unloaded pistol, ready to accept the ammunition. When cleaved to the

mature form, caspases become loaded, ready to fire, but like a pistol, require that additional steps be taken to make the mechanism run. This is analogous to the mature state of caspases, where the loops are almost ready, but not in a catalytically competent conformation. The multiple steps involved, along with the placement of the loops ensure that caspases are not inadvertently initiated, a process which would be very detrimental to a cell.

Because of this large rearrangement of loops that compose the active site, the active site is naturally more flexible than the core of the protein. At first blush, the inherent flexibility of an active site can be more of a hindrance than a benefit for a protein as critical for cell homeostasis as a caspase. It is generally agreed that catalytic sites occur at the interfaces of domains or intersubunits[128], however, catalysis of reactions by proteins requires very specific orientation of the residues involved. The additional support afforded by the surrounding structures in proteins at interfaces can rigidify the active site to allow for proper geometry to undergo catalysis. Additionally, an interface can present a solvent accessible region for the interaction of substrate and enzyme, facilitating the capture of substrate and release of product. In the case of HIV-1 protease, the dimer interface houses the catalytic residues which remain stationary, while two 'flaps' undergo conformation changes that move 7Å during substrate binding[129,130]. This relative burial of the active site as compared to caspases does not affect the enzyme efficiency. HIV-1 protease exhibits a higher turnover of 42.3 s^{-1} while caspases in general exhibit much slower turnover on the order of 1 s^{-1}, even though the active site is much more exposed. However, the K_m of the buried active site in HIV-1 is much lower (3.7 μM) as compared to caspases, which have K_m constants on the order of 60-100 μM. One might assume that

the protein with more flexible loops would have a lower K_m, as the active site binding groove is more flexible and thus less likely to hold substrate. However as we see in this example, this is not the case. The high K_m may be related to the necessity for higher specificity in caspases, combined with the rearrangement of the loops between at least two states in solution.

The discovery of the two distinct caspase-7 states in solution, as described earlier in the pistol metaphor, has wide ranging implications for the further use of caspases. Because the global folds of caspases are so similar, this lends credence to the idea that the two state nature of caspase-7 is not unique among caspases. Along these lines, the large thermal stability increase noted in caspases, in particular caspase-7, can be used in drug design efforts not currently being undertaken. Using methods like thermo sensitive dyes to detect thermal denaturation is established, and has been used in high throughput applications for drug discovery[131]. This method relies on a difference in T_m upon the binding of ligand to differentiate specific and non-specific binders. The noted changes in thermal stability upon binding of ligand vary in caspases. For instance, caspase-7 and caspase-9 share the highest known difference between bound and unbound states of 17 degrees (Caspase-9 thermal melt data generously provided by Kristen Huber, unpublished). This is compared to caspase-3 which has a T_m of greater than 7 degrees, and -6, which has a ΔT_m of only 3 degrees[27]. The low ΔT_m for caspase-6 has can be explained in the appearance of an extended helix that preselects caspase-6 to a more active state, however the lower ΔT_m for caspase-3 relative to caspase-7 is likely due to the inability to measure melting temperatures above 90 degrees on the CD spectrometer used. The similar structures of other caspases suggest that a one would find increase of stability

on the order of 17 degrees upon binding substrate. The energy associated with binding and stabilization on this scale is not trivial. This stabilization in caspases can be compared to the binding of biotin to streptavidin, an incredibly tight interaction, which results in a thermal stabilization of 37 degrees in T_m [132]. In the interaction of streptavidin to biotin[130], the strong binding of the biotin to the hydrogen bonding network, Van der Waals surface burial and high shape complementarity all contribute to the strength of this non-covalent interaction.

In the thermal denaturation studies of caspases to date, a covalent inhibitor has been present in the active site. The covalent nature of this interaction could affect the final active site binding groove arrangement, as these loops could relax and adopt a low energy state that is not necessarily biologically relevant. However, recent studies have shown that the structure of caspase-7 in the presence of a non-covalently bound inhibitor adopts a nearly identical conformation to that of the covalent inhibitor (structural information provided by Scott Eron, unpublished data). This strongly suggests that not only are the peptide-protein contacts sufficient to maintain caspase-7 in an active state, but that the covalent bond at the catalytic cysteine to the inhibitor is not the driving force behind the formation of the active site, but instead the peptide recognition is. From this, we expect that all caspases, when interacting with the target substrate during a recognition period, would adopt a conformation that is significantly stabilized. Thus, the native disulfide we observe in caspase-7 could be stabilizing a non-ideal substrate recognition motif in caspase-7, allowing the substrate to bind non-ideally, but still in a regime that allows for proteolysis.

Designing an inhibitor that targets the loop arrangement and not simply the binding pocket[65], interrupting the formation of the loop bundle buttress can be an effective method for additional caspase control particularly when the loops can be accessed from outside the substrate binding groove. Furthermore, the use of FRET as a measuring tool[133] with labeled residues on L2 and L2' or hydrogen-deuterium exchange would give additional insight to dynamic loop movement in solution not necessarily present in circular dichroism data alone. Structural information can significantly improve the ability to determine the importance of potential disulfide bridges. Stabilization of a particular conformation is a very common method of protein regulation or even activation. In the classic example of chymotrypsinogen, a pair of disulfide bridges stabilizes the active form of the enzyme after the cleavage of residues that active the protein. Design of a redox-controlled protein necessarily relies on the protein adopting a different conformation than the active, this can be done as in caspase-7 where the structure of a known inactive state exists. In the design of a of a redox controlled lysozyme, where only the active conformation was known, Matthews and colleagues engineered in a disulfide that deactivates the protein in the presence of a oxidizing environment[134]. The inherent flexibility of T4 lysozyme allowed the formation of a disulfide bridge that disabled the enzyme by locking the two lobes of the protein together. In our design of a redox-controlled caspase, we used a known structure of caspase-7 in the allosterically inhibited form as a starting point. In the case of caspase-3, the lack of structural information of a L2' loop down conformation precluded the initial structural study of a redox loop locked conformation, even when the native sequence of caspase-3 contains a homologous potential disulfide to R210C in caspase-7. We postulate that because of the structural

similarities within caspases, the design of redox loop locked caspases is not limited to caspases of which we have structures of multiple conformations such as caspase-7. Therefore, the application of a loop lock as we have designed can be integrated into not only caspase-3 and -7, but also into other caspases, which would be eminently useful to allosterically control. As has been previously shown[135,136,109,137], delivery of active proteins into cells is possible. The possibility of delivering a functional caspase into cells with down regulated apoptosis using the redox-controlled method would allow direct and controlled initiation of apoptosis. Delivering caspases directly to where they are needed would be of great benefit for targeted cancer treatments, as minimization of off target effects would limit damage to surrounding tissue.

Additionally, the capability to lock the loops over the caspase allosteric site allows less than optimal designs from computational binding site design to be studied crystallographically in the presence of the inhibitor. For instance, if the inhibitor has poor binding affinity for the new designed allosteric site, a designed caspase can be incubated in the presence of the non-covalent effector molecule in reducing conditions. The small molecule can then be 'trapped' by the locking of the L2' loops over the allosteric cavity simply by the transference of the protein into an oxidizing environment.

A design method of using computational protein redesign can be combined with methods currently underway in the Hardy lab, where specificity for substrate or allosteric sites are being altered. Using EGAD design techniques along with existing mutants and their effect on specificity, fewer new designs may have to be tested in the lab. Similarly, EGAD designs can be improved by the use of real specificity data being collected in the lab.

Data from the studies discussed in this thesis can be applied generally to other systems as demonstrated and discussed. The importance of the large rearrangement of the loop bundle during catalysis as discussed in chapter 2 has implications not only for computational redesign of the allosteric cavity (Chapter 4), and the design for redox control of caspase-7 (Chapter 3), but on designs of caspase specific inhibitors and further studies into the dynamics of caspases. To design inhibitors for a protein with large dynamic rearrangements, taking into account the loop rearrangement can be advantageous, as foreknowledge of sites not amenable to mutation, such as I213 on L2', one can target it specifically.

A direct application of the knowledge from the study of L2' is a design of a redox activatable caspase-7 that makes proteins eminently more drug like, as these designed proteins can be delivered to systems with down regulated apoptosis (Chapter 3).

The strength of the toolset lies in the execution, as EGAD demonstrates by utilizing the vast and inexpensive computational power available to test and therefore design a novel binding site that can function as an allosteric effector. The combination of a large sampling size of potential variants on the order of 10^{24} with multi-factored scoring (Chapter 4) can significantly decrease the wet lab experimental workload. Combined with the approaches outlined above, a new allosteric binding site can be engineered successfully.

It is the hope of the author that future investigators apply the lessons learned and data gathered to improve the impact of their experiments and science as a whole.

APPENDICIES

A. CASPASE-7 IN THE PRESENCE OF AN UNCLEAVABLE PEPTIDE

Abstract

The ability to design individual caspase specific inhibitors is hindered by the shared specificity of the executioner caspases. However, it is clear from numerous studies that the *in vivo* targets of the executioner caspases vary. Thus, the understanding of the substrate-binding groove, and by extension substrate binding and proteolysis in particular, relies heavily on accurate atomic positioning of the catalytic residues. In this work we present a new approach to study protein function and geometry in the presence of an uncleavable substrate mimic, DEVD|GK. Crystallographic and binding evidence in the presence of DEVD|GK suggests an unexpected binding mode and overall caspase-7 conformation not compatible with the canonical caspase-7 substrate-binding models, which all rely on covalent modification of the active-site cysteine by an electrophilic warhead on a peptide-derived inhibitor.

Introduction

Structures of caspase-7 have been solved numerous times in the presence of a covalently bound substrate in the active site[37,61,88]. However, in each of these structures with the catalytic cysteine bound to the peptide, the geometry of the cysteine-histidine dyad (C186, H144) is incompatible with catalysis. In every case, the cysteine's gamma sulfur is at minimum 5Å from the histidine nitrogen. Similarly, in previously reported structures of caspase-7 with nothing bound to the active site[39], the dyad is not in a catalytic orientation, being 4Å to 6Å apart. At the time this project was undertaken,

crystallographic data from the R210C/C246S caspase-7 variant was not yet obtained, and a structure of caspase-7 in the absence of an inhibitor but in a catalytic orientation was not available.

To best determine the catalytic geometry of caspase-7 during catalysis, a structure of caspase-7 bound to an uncleavable substrate was desired. All commercially available active site inhibitors only included residues on the *N*-terminal side of the cleavage, so no information is available about binding of amino acids on the prime side of the binding groove. It was decided that introducing an uncleavable modified peptide bond between the naturally occurring cleavage site *C*-terminal to the aspartate and a subsequent residue would encourage substrate binding using the intrinsic binding sequence (DEVD) but disallow peptide bond cleavage and thus substrate release. Though caspase-3 and caspase-7 specificity has been extensively probed[138,139] it has been conclusively shown that both proteins share a canonical recognition sequence (DEVD). Recent proteome wide substrate analysis of caspase-3 and caspase-7 show differences in P_1' and P_2' recognition sites. The potential additional benefit of mapping the P_1' and P_2' sites encouraged the use of a substrate sequence specific to caspase-7[140]. A sequence of DEVD linked by an uncleavable modified peptide bond to GK was selected for synthesis and analysis.

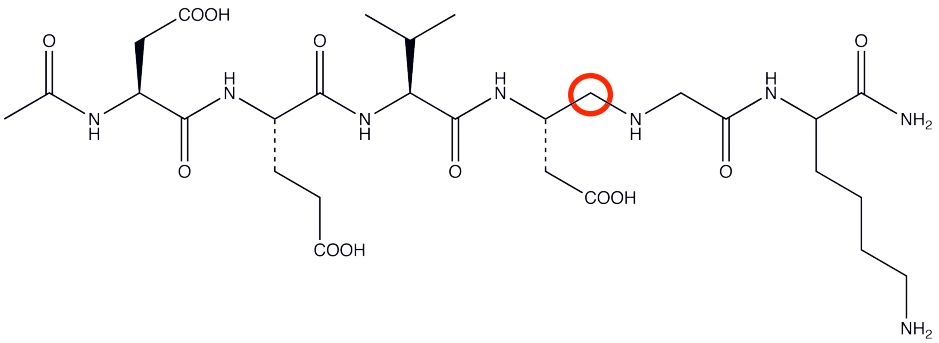


Figure 45: The uncleavable caspase-7 peptide DEVD|GK.

A caspase-7 specific uncleavable peptide consisting of DEVD followed by a modified peptide bond lacking the carbonyl oxygen (red circle) links residues GK.

The synthesis and chemical characterization of DEVD|GK was undertaken by Dr. Sumana Ghosh. Details can be found in the materials and methods of this chapter.

Results

Crystal structure of Caspase-7 with DEVD|GK

After synthesis and characterization, the uncleavable peptide DEVD|GK was co-crystallized with caspase-7 WT protein, and three independent data sets were obtained and solved. As the structure of caspase-7 with DEVD|GK was suspected to be most similar to the structure of caspase-7 in the presence of a covalently bound inhibitor, the model for molecular replacement was 1F1J with loops removed to minimize bias in the loops region. Initial observations of the maps from the molecular replacement solution suggested strongly that a new caspase-7 structure had been observed. When overlaid with 1F1J, density for the loops L2, L3, and L4 were not in an identical conformation or not visible in density entirely. This fact was particularly surprising as all existing structures with substrate bound to the active site had nearly identical loop structures. Even in a structure of caspase-7 with severely impaired activity (see Chapter 2), the loops adopt a

conformation nearly identical to that of WT. Other minor changes across the whole protein were observed during rebuilding.

Only one active site of C7 + DEVD|GK is ordered in the crystal structure. During rebuilding, repeated attempts were made to build in loops near the active site of chain A (one half of the dimer, composed of a large and small subunit), however, this was met with difficulty, as the density remained ambiguous and discontinuous. In contrast, the B chain of the protein not only had better density for the active site loops, but also contained density consistent with the binding of DEVD|GK in a similar motif to the binding of DEVD in 1F1J.

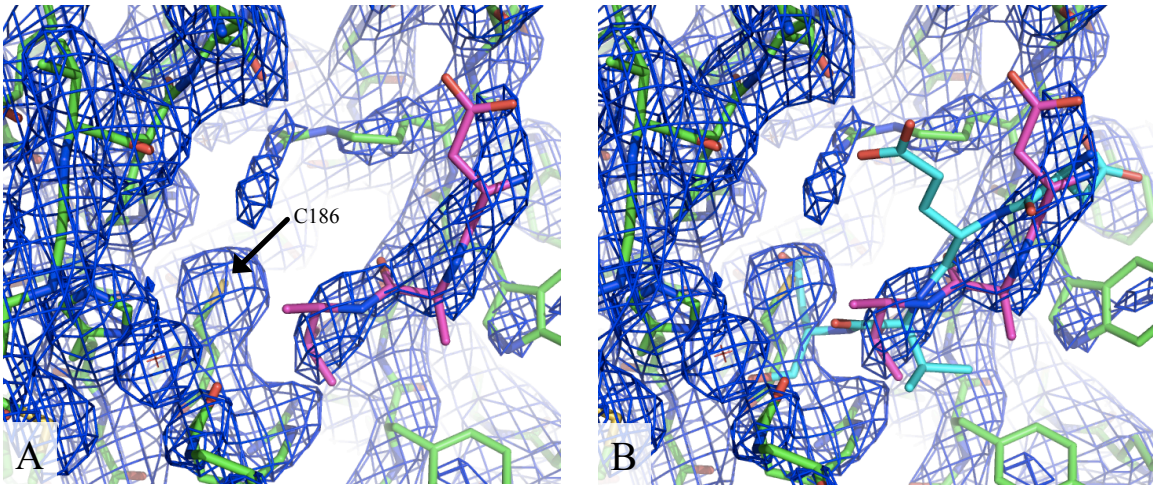


Figure 46: Active site and DEVD visible in one half of dimer

Density for the active site cysteine is only present on one side of the dimer (chain B). (A) Density for the uncleavable peptide is visible for residues DEV (magenta), but is discontinuous and ambiguous further. (B) The uncleavable peptide (magenta) adopts a conformation that is related to, but not identical to the active site bound peptide in 1F1J (cyan).

Attempts to improve density throughout the protein, non-crystallographic symmetry (NCS) across the whole dimer was applied to a refinement step. After this containment was applied, overall refinement statistics did not improve, rather got worse, suggesting that the two halves of the protein could not be treated identically. NCS was successfully

applied to improve refinement statistics in regions of caspase-7 that did not vary from each side. Specifically, the β-sheet regions that form the core of caspases were used. Residues 66-73, 136-141, 176-186, and 218-233 held with tight restraints at the backbone position and medium restraint for the side chains.

Disulfide visible on one side of caspase-7 dimer

Although the binding of the peptide at the active site is somewhat ambiguous, the discovery of disulfide bond formation at residues C100 and C246, between the large and small subunits (Figure 19, Chapter 3) is of great interest. Intriguingly, this disulfide was only visible on one side of the dimer, *opposite* of the side that showed good density for the active site. A map generated with the residues C100 and C246 omitted shows clear and unambiguous contiguous density consistent with a disulfide formation (Figure 47 A). However, the same is not the case on the opposite side of the dimer, where the density for disulfide formation is not evident (Figure 47 B).

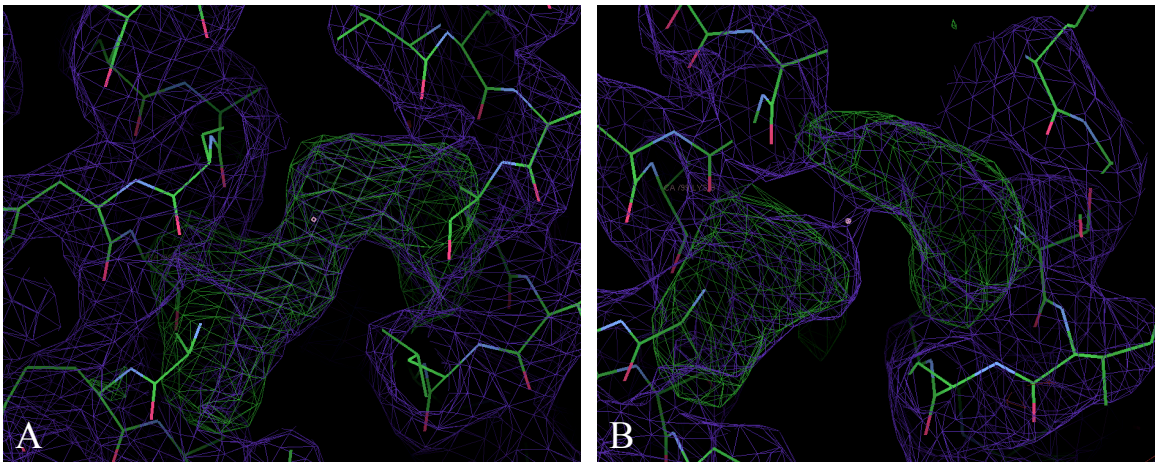


Figure 47: Native C100-C246 disulfide visible in omit map

The formed disulfide is clearly visible on one side of the dimer in a map generated with residues C100 and C246 omitted. (A) Both the F_o-F_c density (green, 3σ contouring) and $2F_o$-F_c (blue, 1σ contouring) is continuous at the site of the omitted C100-C246. (B) On the opposite side of the dimer, the density rendered identically is discontinuous, and does not show evidence of disulfide formation.

We confirmed these results in three independent data sets, each collected on an independent crystal.

C7 + DEVD|GK	Data Set 1	Data Set 2	Data Set 3
Diffraction Data			
Wavelength (Å)	1.00	1.00	1.00
Resolution Range (Å)	50-2.75	50-2.80	50-2.9
Measured reflection	245450	298058	82617
Unique reflections	23376	22243	19670
Completeness (%)	99.9 (100)	100 (100)	99.4 (99.7)
Redundancy	10.5 (10.6)	13.4 (13.4)	4.2 (4.2)
$\langle I/\sigma I \rangle$	38.9 (2.1)	36.0 (2.1)	24.6 (2.2)
Space Group	$P3_221$	$P3_221$	$P3_221$
a = b (Å)	90.028	90.014	90.016
c (Å)	185.261	185.645	185.702
α = β (°)	90	90	90
γ (°)	120	120	120
r_{sym}	7.3 (99.7)	10.4 (89.0)	8.2 (55.6)
Refinement Statistics			
Atoms	3486		
Water molecules	25		
R_{work} (%)	29.6		
R_{free} (%)	23.2		
RMSD bond length (Å)	0.021		
RMSD bond angle (°)	2.15		
Average B-factor (Å^2)	58.3		

Table 7: Crystallographic statistics for three independent datasets of caspase-7 with DEVD|GK

Data sets are stored on the Hardy linux workstation, solitude, in the /home/xray_data/X25 Aug 2010/px10-0295 directory. Data set 1 is stored in the subdirectory /W2_UK_3/5 Data set 2 in /C7_DEVDGK_1_8/ and Data set 3 in /C7_DEVDGK_2_1/

Inhibition of caspase-7 using DEVD and DEVD|GK

In light of the unusual structural results of caspase-7 co-crystallized with DEVD|GK, binding of DEVD|GK vs DEVD was compared. To asses whether DEVD|GK binds in

similar fashion to the native recognition sequence of DEVD, DEVD|GK's inhibition was compared to the peptide DEVD. It was expected that the recognition sequence of DEVD|GK would be sufficiently similar to that of DEVD, causing the inhibition to be similar to product inhibition, and be competitive.

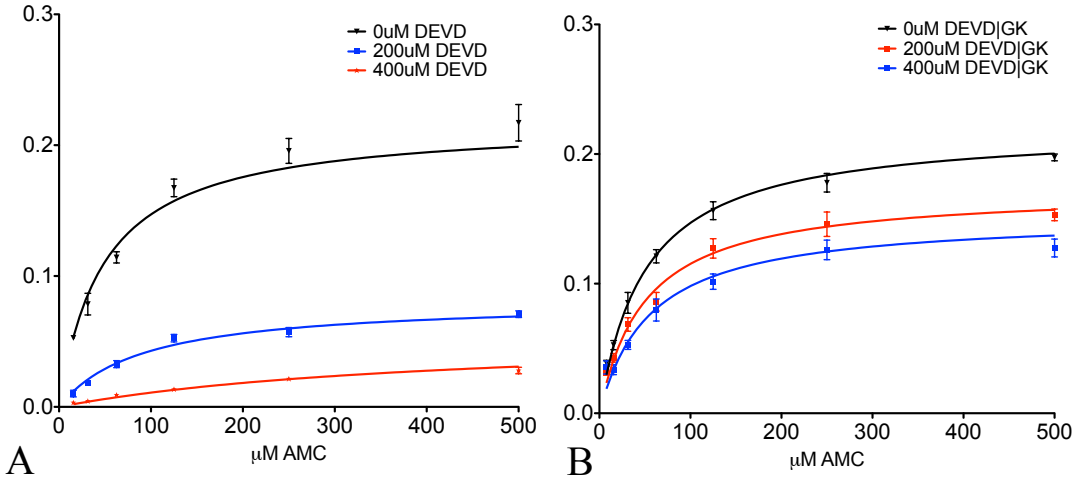


Figure 48: Inhibition curves based on substrate titration

The inhibition titration curves of DEVD (A) vs DEVD|GK (B) measured during a substrate titration using DEVD-AMC as substrate with protein concentrations of 250 nM for DEVD 175 nM for DEVD|GK.

	0 µM		200 µM		400 µM	
	DEVD	**DEVD|GK**	**DEVD**	**DEVD|GK**	**DEVD**	**DEVD|GK**
k_{cat}	217.2	225.2	73.9	179.6	36.0	150.0
K_m	65.3	45.93	95.86	47.8	235.1	46.04

Table 8: Kinetic parameters of caspase-7 in the presence of DEVD and DEVD|GK.

Kinetic parameters are fit from data in Figure 48.

Discussion

The surprising results from the crystal structure, and the unexpected inhibition results suggest a different binding mode for the un-cleavable peptide. In the crystal structure, the peptide is visible to the P2 position, however the density becomes ambiguous and non-contiguous making the placement of the P1 aspartate impossible. Similarly, the P1 pocket is occupied by the catalytic cysteine where in the 1F1J structure the P1 aspartate finds the P1 site (Figure 46). Then minor but significant differences between the 1F1J structure and the structure with the un-cleavable peptide around the active site suggest that DEVD|GK does not bind in the canonically understood fashion.

It is also significant that there are obvious differences between the two halves of the dimer. While the uncleavable peptide can be resolved on one half of the dimer, the opposite side of the dimer is much more disordered. The region surrounding the active site in chain A is missing many more residues. Specifically, the catalytic cysteine in chain A is not visible nor is the uncleavable peptide.

This asymmetry is carried forward in the region of the C100-C246 disulfide. Interestingly, the disulfide is visible only on the side of the dimer where the active site is disordered. On the side of the diner where peptide is visible, the disulfide is clearly not formed (Figure 47). The cause of this asymmetry may be related to the unexpected binding mode of the uncleavable peptide. The uncleavable peptide is not acting like a competitive inhibitor to the fluorescent substrate, rather it is only affecting the k_{cat}, while leaving K_M unaffected thus exhibiting classic noncompetitive inhibition properties. On the other hand, the control DEVD peptide, is exhibiting mixed inhibition in that both K_M

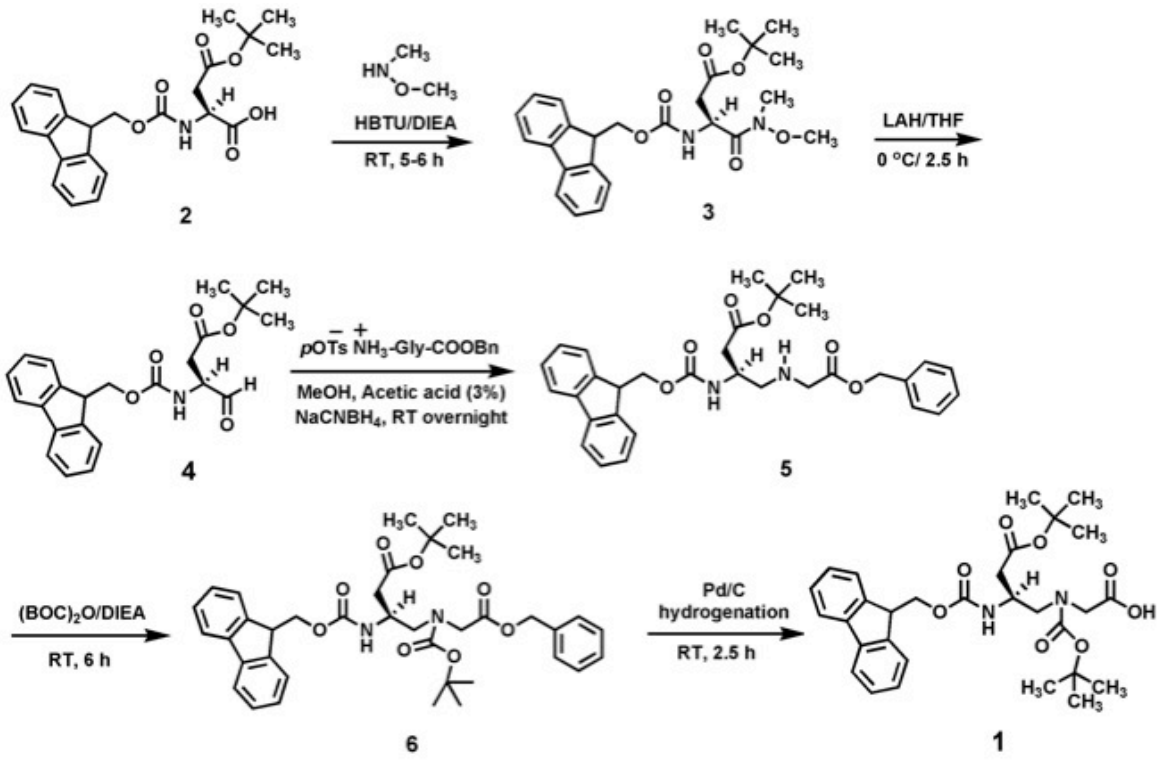
and k_{cat} are being affected. It is therefore difficult to conclude with any certainty on the mechanism of binding or inhibition despite the solution of a crystal structure.

Although the kinetics experiments leave much to be studied, the crystallographic data is more conducive of interpretation. The appearance of resolvable uncleavable peptide on one half of the dimer, while a stabilizing disulfide is formed on the opposite side of the dimer may suggest that caspase-7 can accommodate binding of substrates in odd confirmations. The uncleavable peptide necessarily contains a non-natural peptide bond right at the site of caspase-7 cleavage between D and G in DEVDGK. This un-cleavable bond has sp^3 geometry as opposed to the standard peptide bond sp^2 geometry. One additional conclusion that can be made from the crystallographic data is that the treatment of caspase-7 as a set of identical monomers is not valid. The imposition of non-crystallographic symmetry during structural refinement must be made judiciously, as the dimers appear to act independently.

Materials and Methods

Synthesis of uncleavable bond peptide

The synthesis and characterization of the uncleavable DEVD|GK peptide was done by Dr. Sumana Ghosh during her tenure in the Hardy Lab. Figure used with permission.



Fmoc-L-Asp-N(Me)OMe,(9H-fluoren-9-yl)methyl-(S)-1-(N-methoxy-N-methylcarbonyl)-2 (tert-butoxycarbonyl)ethylcarbamate, (3). N-Fmoc-L-aspartic acid β-tert-butyl ester (2, 2 gm, 4.86 mmol) and N,O-dimethyl hydroxylamine hydrochloride

(0.57 gm, 5.83 mmol) were dissolved in 15 ml DMF. The reaction mixture was cooled and DIEA (2 ml, 11.7 mmol) followed by HBTU (2.2 gm, 5.83 mmol) were added into it. The resulting solution was allowed to stir for 5-6 h at room temperature. At end the reaction mixture was extracted with ethylacetate and successively washed with 0.2 N HCl, saturated NaHCO3 solution and finally with brine. The crude gummy liquid was purified by column chromatography using 0.5 % MeOH/DCM as the eluent. The solvent was evaporated to afford gummy liquid (3, 1.7 gm) as 80% isolated yield. 1H-NMR (CDCl3, 400 MHz) δ (ppm) 1.44 (s, 9H), 2.6 (dd, J1 = 16 Hz, J2 = 7 Hz, 1H), 2.74 (dd, J1 = 16 Hz, J2 = 5.3 Hz, 1H), 3.2 (s, 3H), 3.8 (s, 3H), 4.23 (t, J = 7.1 Hz, 2H), 4.36 (d, J = 7.3 Hz, 2H), 5.04 (m, 1H), 5.72 (d, 1H), 7.32 (t, J = 7.4 Hz, 2H), 7.4 (t, J = 7.4 Hz, 2H), 7.6 (d, J = 7.4 Hz, 2H), 7.76 (d, J = 7.5 Hz, 2H). (Lab Note book, p. 94, 98)

(9H-fluoren-9-yl)methyl-(S)-2-(tert-butoxycarbonyl)-1-formylethylcarbamate,(4). To a solution of Weinreb amide (3, 0.1 gm, 0.22 mmol) in 2 ml anhydrous THF, 13 mg (0.33 mmol) Lithium aluminium hydride was added at 0 oC (ice). The reaction mixture was stirred at 0 oC for 2 h 30 min. and then quenched by addition of 10 ml 5% KHSO4 solution. The resulting solution was extracted with ethylacetate, washed 2 or 3 times with 5 % KHSO4 solution, brine and dried (MgSO4) and concentrated under reduced pressure. The resulting aldehyde was purified by silica column chromatography eluting with DCM and MeOH/DCM mixture. The final compound was obtained using 2-2.5 % MeOH/DCM as eluent . The aldehyde was obtained as gummy liquid almost 70 % (0.06 gm) isolated yield. 1H-NMR (CDCl3, 400 MHz) δ (ppm) 1.45 (s, 9H), 2.78 (dd, J1 = 17.2 Hz, J2 = 4.6 Hz, 1H), 2.96 (dd, J1 = 17.2 Hz, J2 = 4.8 Hz, 1H), 4.24 (t, J = 6.8 Hz, 2H), 4.38-4.46 (m,

3H), 5.9 (d, 1H), 7.32 (t, J = 7.4 Hz, 2H), 7.4 (t, J = 7.5 Hz, 2H), 7.6 (d, J = 7.4 Hz, 2H), 7.77 (d, J = 7.5 Hz, 2H) 9.65 (s, 1H).

(9H-fluoren-9-yl)methyl(S)-1-(((benzyloxy)carbonyl)methylamino)-3-(tert-butoxycarbonyl) propan-2-ylcarbamate, (5). Compound 4 (1.1 gm, 2.8 mmol) was dissolved in 7 ml MeOH. Into that p-toluenesulfonate salt of Gly-OBn (1.13 gm, 3.36 mmol) along with 0.208 ml (3.64 mmol) acetic acid and some amount of molecular sieves were added. The resultant mixture was stirred at room temperature for an hour. Then sodium cyanoborohydride (3.36 mmol, 0.25 gm equivalent to 4 ml of 1 M NaCNBH4 solution in THF) was added dropwise into the reaction mixture at room temperature. Immediately a white precipitate starts forming and continued stirring at room temperature for overnight. The resultant solution was then diluted with methanol, filtered and evaporated. The resultant crude compound was dissolved in ethylacetate and successively washed with water, saturated NaHCO3 solution, brine and dried over Na2SO4. The product was purified using EtOAC/Hexane as eluent and the product was obtained using 20% ethylacetate/hexane as eluent. 1H-NMR (CDCl3, 400 MHz) δ (ppm) 1.45 (s, 9H), 2.51 (dd, 1H), 2.54 (dd, 1H), 2.76 (dd, 1H), 2.82 (dd, 1H), 3.47 (d, J = 6.0 Hz, 2H), 4.05 (m, 1H), 4.23 (t, J = 7 Hz, 2H), 4.4 (d, J = 6.6 Hz, 2H), 5.2 (s, 2H), 5.6 (d, 1H), 7.3-7.4 (m, 9H), 7.6 (d, J = 7.4 Hz, 2H), 7.76 (d, J = 7.5 Hz, 2H).

Synthesis of (6). Compound 5 (0.22 gm, 0.4 mmol) was dissolved in 10 ml anhydrous THF followed by the addition of 14 µl (0.08 mmol) DIEA. The reaction mixture was cooled to 0oC and BOC2O(DiBoc Anhydride) (0.096 gm, 0.44 mmol) was added

dropwise into it. The resultant solution was stirred at room temperature for 6 h. The solvent was evaporated and purified by silica column chromatography. The product was obtained while eluting with 1% MeOH/DCM for an 80% (0.17 gm) isolated yield. 1H-NMR (CDCl3, 400 MHz) δ (ppm) 1.42 (bs, 18H), 2.47 (dd, J1 = 17.0 Hz, J2 = 6.4 Hz, 1H), 2.62 (dd, J1 = 14.0 Hz, J2 = 5.1 Hz, 1H), 3.3-3.4 (m, 1H), 3.5-3.62 (m, 1H), 3.96 (q, J1 = 12.0 Hz, 2H), 4.03-4.05 (m, 1H), 4.2 (t, J = 7 Hz, 2H), 4.35 (dd, J1 = 25.4 Hz, J1 = 5.6 Hz, 2H), 5.2 (s, 2H), 5.6 (d, 1H), 7.3-7.4 (m, 9H), 7.6 (d, J = 7.4 Hz, 2H), 7.74 (d, J = 7.5 Hz, 2H).

Synthesis of (1). Compound 6 (1.2 gm, 1.86 mmol) was dissolved in 20 ml MeOH and Pd/C (Paladium charcoal) (10%, 150 mg) was added into it under inert atmosphere. The reaction mixture was stirred and hydrogenated over 2h 30 min. The resultant solution was filtered and purified by silica column chromatography. The product was obtained while eluting with 2% MeOH/DCM mixture. The yield of the reaction is almost 85%. 1H-NMR (CDCl3, 400 MHz) δ (ppm) 1.42 (bs, 18H), 2.45-2.5 (dd, J1 = 17.0 Hz, J2 = 6.4 Hz, 1H), 2.59-2.64 (dd, J1 = 14.0 Hz, J2 = 5.1 Hz, 1H), 3.3-3.4 (m, 1H), 3.51-3.62 (m, 1H), 3.94 (q, J1 = 18.0 Hz, 2H), 4.03-4.05 (m, 1H), 4.2 (t, J = 7 Hz, 2H), 4.35 (dd, J1 = 25.4 Hz, J1 = 7 Hz, 2H), 5.97 (d, 1H), 7.3 (t, J = 7.4 Hz, 2H), 7.38 (t, J = 7.4 Hz, 2H), 7.6 (m, 2H), 7.74 (d, J = 7.5 Hz, 1H)

Inhibition assays with DEVD and DEVD|GK

Enzyme concentrations were determined by active site titration with covalent substrate DEVD-CHO (*N*-Acetyl-Asp-Glu-Val-Asp-Aldehyde, Enzo Lifesciences) in caspase

activity buffer containing 100 mM HEPES (4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid) pH 7.0, 10% polyethylene glycol 400, 0.1% CHAPS, 5 mM β-mercaptoethanol, and 5 mM $CaCl_2$ using the flourogenic substrate, DEVD-AMC (*N*-acetyl-Asp-Glu-Val-Asp (7-amino-4-methylcoumarin), Enzo Lifesciences), Ex365/Em495. Active site titration samples were incubated over a period of 2 hours in 120 mM NaCl 20 mM Tris pH 8.0 at nanomolar concentrations. Optimal labeling was observed when protein was added to DEVD-CHO solvated in DMSO in 96-well V-bottom plates, sealed with tape, and incubated at room temperature. 22.5 μL and 90 μL aliquots were transferred to black-well plates in duplicate, and assayed with 50-fold molar excess of substrate. For substrate inhibition measurements, 100 nM protein was assayed with DEVD (Biomatic Corporation) or DEVD|GK (synthesized in house by Dr. Sumana Ghosh) present using a fluorescent substrate titration over the range of 0-1000 μM DEVD-AMC over the course of seven minutes. Data for concentrations above 1000 μM was omitted due to observed product inhibition. Assays were performed at 37°C in 25 μL and 100 μL volumes in 96 or 384 well microplate format using a Molecular Devices Spectramax M5 spectrophotometer. Initial velocities versus substrate concentration were fit using Prism software (Graphpad Software) to determine parameters K_m, K_{Iapp} and k_{cat}.

Caspase-7 + DEVD|GK crystallization and x-ray data collection

To prepare caspase-7 with DEVD|GK crystals, protein in a buffer containing 120 mM NaCl and 20 mM Tris pH 8.0 was concentrated using Millipore Ultrafree 5K NMWL membrane concentrators (Millipore) to 7.8 mg/mL as assessed by absorbance at 280 nm. DEVD|GK peptide dissolved in water was added in 4 fold molar excess and incubated for

one hour. Crystal trays were setup at room temperature and grown in 3.5 μL hanging drop with mother liquor consisting of 100 mM citrate buffer, 1M LiSO4 1M NaCl in a 2.5:1 ratio of protein to mother liquor. Crystals grew to a maximum of 250 μm in 7 days at 20 degrees. Crystals were cryoprotected in 20% ethylene glycol in mother liquor with a 60 second incubation, then frozen by rapid immersion in liquid N_2. Data were collected using synchrotron radiation at Brookhaven X25 over 300 frames with 1° oscillations and showed diffraction data to 2.75Å, with frames 1-50, 142, and 175-245 omitted for data set 1. Data set 2 was collected identically over 275 frames, with frames 105-145 omitted. Omitted frames were due to high linear R-factor. Data set 3 was collected identically to data sets 1 and 2 over 70 frames. Indexing, integration and scaling were carried out with HKL2000[73]. Data for the lower resolution crystals were collected during the same data collection trip.

Structure determination

Phase information was obtained by molecular replacement using 1F1J as the search model with residues corresponding to L2' omitted during searches with PHASER[115]. Removed residues were rebuilt into unambiguous density using COOT[116], then refined using Refmac[117] restrained refinement with TLS enabled and NCS applied to residues 66-73, 136-141, 176-186, 233-218. Water and sulfate molecules were inserted manually and checked for stereochemical appropriateness. The final refined model from data set 1 contains residues 55–185, 214–226, 236–304 for chain A and 54–188, 212– 306 for chain B. Uncleavable peptide is modeled for residues DEV as chain E, 25 water molecules are assigned chain O, and 2 sulfates assigned chain S.

B: EGAD SCRIPTS

Outputs and tabulated output data in excel format can be found on the Hardy linux workstation solitude in the /home/shared/EGAD directory.

1SHJ_A.input

```
START

#### Input files ####
# Each of these lines declares the location of an input file
# for the design calculation.

# The PDB file that serves as the template for this monomer
TEMPLATE_PDB
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/1SHJ_A.pdb

# The resparam file that defines prototype residues
RESPARAM_FILE
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/common_inputs/vdw0.90.r
esparam

# The forecfield file that defines atom-specific information,
# such as VDW radii
FORCEFIELD
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/common_inputs/atoms.for
cefield

# The rotamer file that defines the rotamer library
ROTAMER_LIBRARY
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/common_inputs/bbind02_m
ay.rotamers

# The torsion parameter file that defines the torsional correction
# for 1,4 bonded atoms
TORSION_FILE
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/common_inputs/opls.tors
ions

# The SASA points file that defines test points on the unit sphere
# for calculation of SASA.
SASA_POINTS_FILE
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/common_inputs/sasa.txt


#### Energy table file settings ####
# These two parameters instruct the application where to find
# pre-calculated energy tables, and whether to load them or calculate
# them fresh. Calculating fresh tables is recommended, in general.
ENERGY_TABLE
      /mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/tet_caspase
LOAD_ENERGIES     0
```

```
#### Variable position definitions ####
# This section defines which positions will be allowed to mutate
# during the design calculation. Its a simple table of position indices
# (from the PDB file) and the residues allowed at that position.
#
# Any positions not mentioned here are assumed to be "fixed", that is
# to say they will not move during a design. Any positions defined to
# be the same as their wild-type counterpart will be allowed to change
# rotamers.
#
# The special keyword "all" will allow any residue type (except GLY,
# CYS, and PRO).
VARIABLE_POSITIONS
    159         all  # ILE
    211         all  # Tyr
    213         all  # Ile
    221         all  # PHE
    223         TYR  # ONLY ROTOMER CHANGE
    290         all  # CYS
    292         all  # VAL
END

#### Energy function settings ####
# Each of these flags sets whether a portion of the energy function
# should be used in the design calculation. A '1' means that it will
# be included, while a '0' means it will not.
VDW_FLAG          1
COULOMB_FLAG      1
TORSION_FLAG      1
SASA_FLAG         1
GB_FLAG           1
HBOND_FLAG        0
REFERENCE_FLAG    1

#### Energy function parameters ####
# Various parameters that modify how the energy functions behave.
# At this point, I just use this to define what scaling factor to
# apply to the VDW radius.
VDW_RADII_SCALE   0.90

END
```

1SHJ_B.input

```
START

TEMPLATE_PDB
/mnt/nmr2/people/witold/Hardy_collab/C7_Tet_1.0/1SHJ_B.pdb

VARIABLE_POSITIONS
    159         all  # ILE
    211         all  # Tyr
    213         all  # Ile
    221         all  # PHE
    223         TYR  # ONLY ROTOMER CHANGE
    290         all  # CYS
    292         all  # VAL
END

END
```

Dox.liginput

```
START

PARAMETER_FILE          dox.ligparam
LOCATION
    COORDINATES           53.538    13.249     3.811
    SPREAD_RADIUS       2.0
    SPREAD_GRAIN        3.0
    XYZ_ROTS            3 3 3
    XYZ_RANGE           0.195 0.195 0.195
END_LOCATION

END
```

Dox.ligparm

```
EGAD Ligand Parameter file for Doxycycline
by Arnab Chowdry; edited by Witold Witkowski
8-14-2007

Format Key:
  Each LIGAND entry holds the information necessary to define one
ligand.

  CORD lines hold atom coordinate information:
        Atom Number Atom Name   Atom Type  Rigid Body     x       y
z    Charge
  CORD       1          CA          CT         3        2.340   1.155
10.616  0.000

  CNCT lines hold atom connectivity information, expressed as a list of
```

```
      atoms directly connected to the first one:
       Parent Atom     Bonded Atoms
  CNCT       1            2 4 5 6

  ROTB lines hold rotable bond information. Atoms connected to the
first atom (excluding the second atom) will be dependent on the rotable
bond:
        First Atom  Second Atom
  ROTB
START
LIGAND DOX
CORD  1     C     CA    0       52.569  15.775   3.409      0.084104
CORD  2     C     CA    0       53.435  16.652   2.849      0.434036
CORD  3     C     CA    0       54.774  16.229   3.364       0.007574
CORD  4     C     CA    0       54.924  14.645   3.172       0.060912
CORD  5     C     CA    0       53.683  13.942   2.473       0.672938
CORD  6     C     CA    0       52.192  14.519   2.676        -1.551142
CORD  7     C     CA    0       53.795  12.445   2.762        -1.127784
CORD  8     C     CA    0       53.736  12.115   4.245        -0.567027
CORD  9     C     CA    0       54.748  12.933   5.031        0.853653
CORD  10    C     CA    0       55.347  14.009   4.492        0.212514
CORD  11    C     CA    0       55.006  12.507   6.420      0.493023
CORD  12    C     CA    0       54.390  11.248   6.895      0.557629
CORD  13    C     CA    0       53.832  10.334   5.986   -0.23334
CORD  14    C     CA    0       53.943  10.603   4.502   -0.57567
CORD  15    C     CA    0       53.224   9.178   6.445      -0.333433
CORD  16    C     CA    0       53.173   8.912   7.812      -0.409527
CORD  17    C     CA    0       53.732   9.788   8.720      -0.022899
CORD  18    C     CA    0       54.348  10.950   8.269      -0.073312
CORD  19    C     CH3   0       50.154  15.356   1.682      -0.405255
CORD  20    C     CN    0       53.503  16.985   1.411      0.573606
CORD  21    C     CH3   0       51.536  13.926   0.302   -0.200686
CORD  22    C     CH3   0       55.273  10.044   3.956   -0.758613
CORD  23    H     HA    0       53.713   9.601   9.774      0.264047
CORD  24    H     HO    0       55.356  12.506   8.863      0.584718
CORD  25    H     HO    0       52.964  11.257   1.338   0.499983
CORD  26    H     HC    0       52.545  13.671   0.028   0.176319
CORD  27    H     HC    0       51.097  14.439  -0.543   0.194366
CORD  28    H     HC    0       50.980  13.010   0.486      0.202262
CORD  29    H     HC    0       49.764  15.781   0.768      0.193855
CORD  30    H     HC    0       49.470  14.575   2.019      0.180426
CORD  31    H     HC    0       50.173  16.138   2.430      0.446138
CORD  32    H     HN    0       51.968  18.326   1.517   0.291143
CORD  33    H     HA    0       53.882  14.106   1.427      0.243309
CORD  34    H     HA    0       51.607  13.843   3.297   0.266719
CORD  35    H     HA    0       54.743  12.117   2.358      0.246798
CORD  36    H     HA    0       52.723  12.328   4.589   0.255958
CORD  37    H     HA    0       53.130  10.087   4.007   0.239156
CORD  38    H     HA    0       52.795   8.485   5.745      0.241564
CORD  39    H     HA    0       52.698   8.017   8.163      0.433676
CORD  40    H     HN    0       52.664  18.115  -0.034      0.554144
CORD  41    H     HO    0       52.374  14.855   5.164   0.569422
CORD  42    H     HO    0       56.723  15.392   4.666   0.212854
CORD  43    H     HC    0       55.387   9.012   4.261   0.225906
CORD  44    H     HC    0       56.126  10.599   4.327   0.173835
CORD  45    H     HC    0       55.301  10.066   2.873   0.542809
CORD  46    H     HO    0       56.012  15.052   1.540      0.447089
```

```
CORD  47    N     N     0       51.503  14.842   1.440      -0.873187
CORD  48    N     N     0       52.619  17.869   0.928      -0.51929
CORD  49    O     OH    0       52.741  15.627   4.741   -0.591843
CORD  50    O     OH    0       56.387  14.606   5.091   -0.5372
CORD  51    O     OH    0       56.045  14.464   2.297   0.201514
CORD  52    O     O     0       55.639  13.205   7.222      -0.611946
CORD  53    O     OH    0       54.879  11.754   9.222      -0.765921
CORD  54    O     OH    0       52.718  11.745   2.115   -0.553176
CORD  55    O     O     0       54.419  16.517   0.708   -0.625366
CORD  56    O     O     0       55.718  16.888   3.715   -0.501389
CNCT   1    2    6   49
CNCT   2    1    3   20
CNCT   3    2    4   56
CNCT   4    3    5   10   51
CNCT   5    4    6    7   33
CNCT   6    1    5   34   47
CNCT   7    5    8   35   54
CNCT   8    7    9   14   36
CNCT   9    8   10   11
CNCT  10    4    9   50
CNCT  11    9   12   52
CNCT  12   11   13   18
CNCT  13   12   14   15
CNCT  14    8   13   22   37
CNCT  15   13   16   38
CNCT  16   15   17   39
CNCT  17   16   18   23
CNCT  18   12   17   53
CNCT  19   29   30   31   47
CNCT  20    2   48   55
CNCT  21   26   27   28   47
CNCT  22   14   43   44   45
CNCT  23   17
CNCT  24   53
CNCT  25   54
CNCT  26   21
CNCT  27   21
CNCT  28   21
CNCT  29   19
CNCT  30   19
CNCT  31   19
CNCT  32   48
CNCT  33    5
CNCT  34    6
CNCT  35    7
CNCT  36    8
CNCT  37   14
CNCT  38   15
CNCT  39   16
CNCT  40   48
CNCT  41   49
CNCT  42   50
CNCT  43   22
CNCT  44   22
CNCT  45   22
CNCT  46   51
CNCT  47    6   19   21
```

```
CNCT  48   20   32   40
CNCT  49    1   41
CNCT  50   10   42
CNCT  51    4   46
CNCT  52   11
CNCT  53   18   24
CNCT  54    7   25
CNCT  55   20
CNCT  56    3
done
STOP
```

Etable.cmdline

```
allow-wt
egad-input  1SHJ_A.input,1SHJ_B.input
ligand-inputs     dox.liginput
```

Machines

```
michael:2
gabriel:2
raphael:2
uziel:2
sariel:2
```

Monte_carlo.cmdline

```
random-seed 116200268
num-runs    100000
num-saved   1000
num-steps   10000
etable      dox_caspase.etable
groups      dox_caspase.groups
```

Montecarlo.sc

```
#!/bin/bash
# Monte Carlo Minimizer Script, by Witold

echo "Running Minimizer"
echo "Monte Carlo Simulation"
echo monte_carlo.cmdline
mpirun -all-cpus -machinefile machines ~/EGAD_Clients/bin/MPIMonteCarlo
--CLPLoad=monte_carlo.cmdline &> monte_carlo.log &

echo "The script should be running, checking monte_carlo.log for
results"
echo "Push CTRL+C to quit log file"
tail -f monte_carlo.log
```

Movestuff.sc

```
#This copies the important files to a new dir.
#Edit the last bit!!!!!
#Written by Witold, 6-6-07

cp -r ./common_inputs 1SHJ_* dox.* machines *.cmdline *.py *.sc *.pdb
*.pml ../doxdes3.2.27
# rm solu*.*
echo "That Should do it!"
echo "Now remember to edit this file before you run it again!"
```

Presurf.sc

```
clear
echo -e "         \E[5mI'm about to delete all the previous surfnet
files\E[25m"
echo "                              Push control-c to stop me!"
sleep 1
echo "5"
sleep 1
echo "4"
sleep 1
echo "3"
sleep 1
echo "2"
sleep 1
echo "1"
rm *.srf
rm *.grf
rm surfnet.ps
rm gaps*
rm volumes*
rm *.out
clear
echo -e "                                     \E[5mRunning Surfnet...\E[25m"
surfnet &> surfnet.out
sleep 7
surface &> surface.out
sleep 5
surfplot &> surfplot.out
sleep 10
clear
echo "                                                All done"
```

Print_structuresnue.py

```
#!/usr/bin/python
#
# This script parses out the result data from a run of MPIMonteCarlo
# and generates command lines for use with MPIPolymerTable. It has
# been modified from previous versions to accept two lists of input
# files: one for proteins and another for ligands.
#
# It also searches for duplicate solutions and removes them from the
# output.
#
import re,sys,os
```

```
if (len(sys.argv) < 3):
    print "Please specify a data file to process and associated input
files."
    sys.exit(1)
# Maps sequence data to energy
SOLUTIONS = {}
# Maps energy value to 3-letter code sequence
ENERGY_TO_SEQ = {}
# The first file on the command line is an output file from a run of
# MPIMonteCarlo, and should contain result lines that we need to parse
# and store.
find_data = re.compile("^OUTPUT:\s+\[[\d]+\][\s]+([\d\-
\.]+)\s+((?:[\d]+\.)+)")
find_seq = re.compile("^OUTPUT:\s+\[[\d]+\][\s]+([\d\-\.]+)\s+((?:[A-
Z]{3}\.)+)")
find_point = re.compile("^ENERGY:\s+([a-zA-Z\s]*) Point Energy is
([\d\-\.]+)")
fp = open(sys.argv[1],'r')
for line in fp.readlines():
    match = find_data.match(line)
    if match:
        SOLUTIONS[match.group(2).replace('.',',')] =
float(match.group(1))
    else:
        match = find_seq.match(line)
        if match:
            ENERGY_TO_SEQ[float(match.group(1))] = match.group(2)
fp.close()
# Now we want to sort by energy value. To do this, we reverse the
dictionary
# into a sorted array.
sorted = [(v,k) for k,v in SOLUTIONS.items()]
sorted.sort()
# Open a file to print summary data for the structures generated.
fp = open("print_summary.txt",'w')
i = 0
energy_cats = {}
energy_keys = []
for energy,sol in sorted:
    command = "MPIPolymerTable -i " + sys.argv[2] + " -w -p"
    sol = sol[:-1]
    if (len(sys.argv) > 3):
        command += " -l " + sys.argv[3]
    else:
      split_string = sol.split(",")
      sol = ",".join(split_string[:-1])
    command += " -c " + sol + " -o solution_" + str(i) + ".pdb"
    output = os.popen(command, 'r')
    for line in output.readlines():
        match = find_point.match(line)
        if match:
            energy_cats[match.group(1)] = match.group(2)
    if i == 0:
        # Print headers
        fp.write("SOLUTION\tENERGY\tSEQUENCE")
        energy_keys = energy_cats.keys()
        for key in energy_keys:
```

```
            fp.write("\t"+key)
        fp.write("\n")
    fp.write("solution_" + str(i) +
":\t"+str(energy)+"\t"+ENERGY_TO_SEQ[energy])
    for key in energy_keys:
        fp.write("\t"+energy_cats[key])
        energy_cats[key] = "N/A" # reset for next round
    fp.write("\n")
    i += 1
    if i == 100: break
# Close the summary file
fp.close()
```

Struct.sc

```
#!/bin/bash
# Structure Compilation script, By Witold (edited for +/- ligand 5-16-
08)
# Uses the pthyon script, print_structures.py

echo "Finding Structures!"
python print_structuresnue.py monte_carlo.log 1SHJ_A.input,1SHJ_B.input
echo "Renaming output file to summary_-_ligand.txt"
mv print_summary.txt summary_-_ligand.txt
python print_structuresnue.py monte_carlo.log 1SHJ_A.input,1SHJ_B.input
dox.liginput
echo "Renaming output file to summary_+_ligand.txt"
mv print_summary.txt summary_+_ligand.txt
echo "DONE"
```

Tabling.sc

```
#!/bin/bash
# Tabling Script, by Witold

rm etable.log

echo "Running Tabling"
mpirun -all-cpus -machinefile machines
~/EGAD_Clients/bin/MPIPolymerTable --CLPLoad=etable.cmdline &>
etable.log &

echo "The script should be running, I'm printing the log file as it is
written."
echo "push CTRL+C to escape"

tail -f etable.log
```

Top5.pml

```
load solution_0.pdb
load solution_1.pdb
load solution_2.pdb
load solution_3.pdb
load solution_4.pdb
```

```
# The interesting bits happen here:
remove hydrogen
#load the mutated residues
sele
/solution_0///147+159+214+215+216+217+221+223+290+292+294+209+210+211+2
12+148
create Sol_0Mut, sele
delete sele
sele
/solution_1///147+159+214+215+216+217+221+223+290+292+294+209+210+211+2
12+148
create Sol_1Mut, sele
delete sele
sele
/solution_2///147+159+214+215+216+217+221+223+290+292+294+209+210+211+2
12+148
create Sol_2Mut, sele
delete sele
sele
/solution_3///147+159+214+215+216+217+221+223+290+292+294+209+210+211+2
12+148
create Sol_3Mut, sele
delete sele
sele
/solution_4///147+159+214+215+216+217+221+223+290+292+294+209+210+211+2
12+148
create Sol_4Mut, sele
delete sele
#load and select interesting residues
#sele ////187+148+160+167+225
#create Pocket, sele
#delete sele
show surface
hide lines
delete sele
create Sol_0Dox, /solution_0///DOX
create Sol_1Dox, /solution_1///DOX
create Sol_2Dox, /solution_2///DOX
create Sol_3Dox, /solution_3///DOX
create Sol_4Dox, /solution_4///DOX
show spheres, Sol_0Dox
show spheres, Sol_1Dox
show spheres, Sol_2Dox
show spheres, Sol_3Dox
show spheres, Sol_4Dox
show dots, Sol_0Dox
show dots, Sol_1Dox
show dots, Sol_2Dox
show dots, Sol_3Dox
show dots, Sol_4Dox
show sticks, Mutated
#show surface, Pocket
show spheres, DOX
sele ////223
create Switch, sele
color red, Switch
show sticks, Switch
```

```
delete sele
load ../1f1J.pdb
align 1f1j, solution_0
hide lines, 1f1j
hide sticks, 1f1j
show cartoon, 1f1j
delete sele
sele /1f1j///223+523
create 1F1J_Tyr, sele
delete sele
color blue, 1F1J_Tyr
show sticks, 1F1J_Tyr
load dox.pdb
```

BIBLIOGRAPHY

1. Kerr JF, Wyllie AH, Currie AR (1972) Apoptosis: a basic biological phenomenon with wide-ranging implications in tissue kinetics. Br J Cancer 26:239-257.

2. Kuida K, Zheng TS, Na S, Kuan CY, Yang D, Karasuyama H, Rakic P, Flavell RA (1996) Decreased apoptosis in the brain and premature lethality in CPP32-deficient mice. Nature 384:368-372.

3. Milligan C, Prevette D, Yaginuma H, Homma S, Cardwellt C, Fritz L, Tomaselli K, Oppenheim R, Schwartz L (1995) Peptide inhibitors of the ICE protease family arrest programmed cell death of motoneurons in vivo and in vitro. Neuron 15:385-393.

4. Davis EC, Popper P, Gorski RA (1996) The role of apoptosis in sexual differentiation of the rat sexually dimorphic nucleus of the preoptic area. Brain research 734:10-18.

5. Chung WCJ, Swaab DF, De Vries GJ (2000) Apoptosis during sexual differentiation of the bed nucleus of the stria terminalis in the rat brain. Journal of neurobiology 43:234-243.

6. Quadros PS, Pfau JL, Goldstein AYN, De Vries GJ, Wagner CK (2002) Sex differences in progesterone receptor expression: a potential mechanism for estradiol-mediated sexual differentiation. Endocrinology 143:3727.

7. Morris JA, Jordan CL, Breedlove SM (2004) Sexual differentiation of the vertebrate nervous system. Nature Neuroscience 7:1034-1039.

8. Polakowska RR, Piacentini M, Bartlett R, Goldsmith LA, Haake AR (1994) Apoptosis in human skin development: morphogenesis, periderm, and stem cells. American Journal of Anatomy 199:176-188.

9. Milligan CE, Schwartz LM (1997) Programmed cell death during animal development. British medical bulletin 53:570.

10. Levine AJ (1993) The tumor suppressor genes. Annual review of biochemistry 62:623-651.

11. Hoffman B, Liebermann D (1994) Molecular controls of apoptosis: differentiation/growth arrest primary response genes, proto-oncogenes, and tumor suppressor genes as positive & negative modulators. Oncogene 9:1807-1812.

12. Cohen JJ (1993) Apoptosis: the physiologic pathway of cell death. Hospital practice (Office ed) 28:35.

13. Tanaka M, Ito H, Adachi S, Akimoto H, Nishikawa T, Kasajima T, Marumo F, Hiroe M (1994) Hypoxia induces apoptosis with enhanced expression of Fas antigen messenger RNA in cultured neonatal rat cardiomyocytes. Circulation research 75:426.

14. Uyama O, Matsuyama T, Michishita H, Nakamura H, Sugita M (1992) Protective effects of human recombinant superoxide dismutase on transient ischemic injury of CA1 neurons in gerbils. Stroke 23:75.

15. Levine B, Huang Q, Isaacs JT, Reed JC, Griffin DE, Hardwick JM (1993) Conversion of lytic to persistent alphavirus infection by the bcl-2 cellular oncogene.

16. Henkart PA (1994) Lymphocyte-mediated cytotoxicity: two pathways and multiple effector molecules. Immunity 1:343.

17. Yuan J, Shaham S, Ledoux S, Ellis HM, Horvitz HR (1993) The C. elegans cell death gene ced-3 encodes a protein similar to mammalian interleukin-1 [beta]-converting enzyme. Cell 75:641-652.

18. Thornberry NA, Bull HG, Calaycay JR, Chapman KT, Howard AD, Kostura MJ, Miller DK, Molineaux SM, Weidner JR, Aunins J (1992) A novel heterodimeric cysteine protease is required for interleukin-1 processing in monocytes. Nature 356:768-774.

19. Stennicke HR, Salvesen GS (1998) Properties of the caspases. Biochim Biophys Acta 1387:17-31.

20. Stennicke H, Salvesen G (1999) Catalytic properties of the caspases. Cell death and differentiation 6:1054.

21. Denault JB, Salvesen GS (2002) Caspases: keys in the ignition of cell death. Chemical reviews 102:4489-4500.

22. Salvesen GS, Riedl SJ (2007) Caspase mechanisms. Programmed Cell Death in Cancer Progression and Therapy:13-23.

23. Green DR (1998) Apoptotic pathways: the roads to ruin. Cell 94:695-698.

24. Saleh A, Srinivasula SM, Acharya S, Fishel R, Alnemri ES (1999) Cytochrome c and dATP-mediated oligomerization of Apaf-1 is a prerequisite for procaspase-9 activation. Journal of Biological Chemistry 274:17941.

25. Bratton S, Salvesen G (2010) Regulation of the Apaf-1-caspase-9 apoptosome. Journal of cell science 123:3209.

26. Milam SL, Clark AC (2009) Folding and assembly kinetics of procaspase-3. Protein Sci 18:2500-2517.

27. Vaidya S, Velazquez-Delgado EM, Abbruzzese G, Hardy JA (2011) Substrate-Induced Conformational Changes Occur in All Cleaved Forms of Caspase-6. J Mol Biol 406:75-91.

28. Rotonda J, Nicholson DW, Fazil KM, Gallant M, Gareau Y, Labelle M, Peterson EP, Rasper DM, Ruel R, Vaillancourt JP (1996) The three-dimensional structure of apopain/CPP32, a key mediator of apoptosis. Nature Structural & Molecular Biology 3:619-625.

29. Mittl PRE, Di Marco S, Krebs JF, Bai X, Karanewsky DS, Priestle JP, Tomaselli KJ, Gr¸tter MG (1997) Structure of recombinant human CPP32 in complex with the tetrapeptide acetyl-Asp-Val-Ala-Asp fluoromethyl ketone. Journal of Biological Chemistry 272:6539.

30. Riedl SJ, Renatus M, Schwarzenbacher R, Zhou Q, Sun C, Fesik SW, Liddington RC, Salvesen GS (2001) Structural basis for the inhibition of caspase-3 by XIAP. Cell 104:791-800.

31. Ni CZ, Li C, Wu JC, Spada AP, Ely KR (2003) Conformational restrictions in the active site of unliganded human caspase 3. Journal of Molecular Recognition 16:121-124.

32. Feeney B, Pop C, Swartz P, Mattos C, Clark AC (2006) Role of loop bundle hydrogen bonds in the maturation and activity of (Pro) caspase-3. Biochemistry 45:13249-13263.

33. Ganesan R, Mittl PRE, Jelakovic S, Grutter MG (2006) Extended substrate recognition in caspase-3 revealed by high resolution X-ray structure analysis. Journal of molecular biology 359:1378-1388.

34. Walters J, Pop C, Scott FL, Drag M, Swartz P, Mattos C, Salvesen GS, Clark AC (2009) A constitutively active and uninhibitable caspase-3 zymogen efficiently induces apoptosis. Biochemical Journal 424:335.

35. Baumgartner R, Meder G, Briand C, Decock A, D'arcy A, Hassiepen U, Morse R, Renatus M (2009) The crystal structure of caspase-6, a selective effector of axonal degeneration. Biochem J 423:429-439.

36. Wang XJ, Cao Q, Liu X, Wang KT, Mi W, Zhang Y, Li LF, LeBlanc AC, Su XD (2010) Crystal structures of human caspase 6 reveal a new mechanism for intramolecular cleavage self-activation. EMBO reports 11:841-847.

37. Wei Y, Fox T, Chambers SP, Sintchak JA, Coll JT, Golec J, Swenson L, Wilson KP, Charifson PS (2000) The structures of caspases-1,-3,-7 and-8 reveal the basis for substrate and inhibitor selectivity. Chemistry & Biology 7:423-432.

38. Chai J, Shiozaki E, Srinivasula SM, Wu Q, Dataa P, Alnemri ES, Shi Y (2001) Structural basis of caspase-7 inhibition by XIAP. Cell 104:769-780.

39. Chai J, Wu Q, Shiozaki E, Srinivasula SM, Alnemri ES, Shi Y (2001) Crystal Structure of a Procaspase-7 Zymogen:: Mechanisms of Activation and Substrate Binding. Cell 107:399-407.

40. Huang Y, Park YC, Rich RL, Segal D, Myszka DG, Wu H (2001) Structural Basis of Caspase Inhibition by XIAP:: Differential Roles of the Linker versus the BIR Domain. Cell 104:781-790.

41. Riedl SJ, Fuentes-Prior P, Renatus M, Kairies N, Krapp S, Huber R, Salvesen GS, Bode W (2001) Structural basis for the activation of human procaspase-7. Proceedings of the National Academy of Sciences of the United States of America 98:14790.

42. Hardy JA, Lam J, Nguyen JT, O'brien T, Wells JA (2004) Discovery of an allosteric site in the caspases. Proceedings of the National Academy of Sciences of the United States of America 101:12461.

43. Siegel JB, Steinmetz WE, Long GL (1980) A computer-assisted model for estimating protein secondary structure from circular dichroic spectra: comparison of animal lactate dehydrogenases. Analytical Biochemistry 104:160-167.

44. Furie B, Bing D, Feldmann RJ, Robison DJ, Burnier J, Furie BC (1982) Computer-generated models of blood coagulation factor Xa, factor IXa, and thrombin based upon structural homology with other serine proteases. Journal of Biological Chemistry 257:3875.

45. Humblet C, Marshall GR (1981) Three dimensional computer modeling as an aid to drug design. Drug Development Research 1:409-434.

46. Arakawa M, Miyao T, Funatsu K (2010) Systematic Generation of Chemical Structures for Rational Drug Design Based on QSAR Models. Current computer-aided drug design.

47. Jallat S, Carvallo D, Tessier L, Roecklin D, Roitsch C, Ogushi F, Crystal R, Courtney M (1986) Altered specificities of genetically engineered 1 antitrypsin variants. Protein engineering 1:29.

48. Sowdhamini R, Srinivasan N, Shoichet B, Santi DV, Ramakrishnan C, Balaram P (1989) Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. Protein engineering 3:95.

49. Pokala N, Handel TM (2001) Review: protein design--where we were, where we are, where we're going. Journal of Structural Biology 134:269-281.

50. Desjarlais JR, Handel TM (1995) De novo design of the hydrophobic cores of proteins. Protein Science 4:2006-2018.

51. Dahiyat BI, Mayo SL (1997) De novo protein design: Fully automated sequence selection. Science 278:82.

52. Ashworth J, Havranek JJ, Duarte CM, Sussman D, Monnat Jr RJ, Stoddard BL, Baker D (2006) Computational redesign of endonuclease DNA binding and cleavage specificity. Nature 441:656-659.

53. Voigt CA, Gordon DB, Mayo SL (2000) Trading accuracy for speed: A quantitative comparison of search algorithms in protein sequence design. J Mol Biol 299:789-803.

54. Mendes J, Guerois R, Serrano L (2002) Energy estimation in protein design. Current opinion in structural biology 12:441-446.

55. Dahiyat BI (2006) In silico protein design: fitting sequence onto structure. Methods Mol Biol 316:359-374.

56. Simons KT, Bonneau R, Ruczinski I, Baker D (1999) Ah Initio: Prediction Reports-Ah Initio Protein Structure Prediction of CASP III Targets Using ROSETTA. Proteins-Structure Function and Genetics:171-176.

57. Kuhlman B, Dantas G, Ireton GC, Varani G, Stoddard BL, Baker D (2003) Design of a novel globular protein fold with atomic-level accuracy. Science 302:1364-1368.

58. Ashworth J, Taylor GK, Havranek JJ, Quadri SA, Stoddard BL, Baker D (2010) Computational reprogramming of homing endonuclease specificity at multiple adjacent base pairs. Nucleic Acids Research 38:5601.

59. Siegel JB, Zanghellini A, Lovick HM, Kiss G, Lambert AR, St Clair JL, Gallaher JL, Hilvert D, Gelb MH, Stoddard BL (2010) Computational design of an enzyme catalyst for a stereoselective bimolecular Diels-Alder reaction. Science 329:309.

60. Wilson KP, Black JA, Thomson JA, Kim EE, Griffith JP, Navia MA, Murcko MA, Chambers SP, Aldape RA, Raybuck SA, et al. (1994) Structure and mechanism of interleukin-1 beta converting enzyme. Nature 370:270-275.

61. Agniswamy J, Fang B, Weber IT (2007) Plasticity of S2-S4 specificity pockets of executioner caspase-7 revealed by structural and kinetic analysis. FEBS J 274:4752-4765.

62. Berger AB, Witte MD, Denault JB, Sadaghiani AM, Sexton K, Salvesen GS, Bogyo M (2006) Identification of early intermediates of caspase activation using selective inhibitors and activity-based probes. Molecular cell 23:509-521.

63. Denault JB, Bekes M, Scott FL, Sexton KM, Bogyo M, Salvesen GS (2006) Engineered hybrid dimers: tracking the activation pathway of caspase-7. Mol Cell 23:523-533.

64. Roy S, Bayly CI, Gareau Y, Houtzager VM, Kargman S, Keen SLC, Rowland K, Seiden IM, Thornberry NA, Nicholson DW (2001) Maintenance of caspase-3 proenzyme dormancy by an intrinsic ìsafety catchî regulatory tripeptide. Proceedings of the National Academy of Sciences of the United States of America 98:6132.

65. Choong IC, Lew W, Lee D, Pham P, Burdett MT, Lam JW, Wiesmann C, Luong TN, Fahr B, DeLano WL (2002) Identification of potent and selective small-molecule inhibitors of caspase-3 through the use of extended tethering and structure-based drug design. Journal of medicinal chemistry 45:5005-5022.

66. Fu G, Chumanevich AA, Agniswamy J, Fang B, Harrison RW, Weber IT (2008) Structural basis for executioner caspase recognition of P5 position in substrates. Apoptosis 13:1291-1302.

67. Xu J, Baase WA, Baldwin E, Matthews BW (1998) The response of T4 lysozyme to large-to-small substitutions within the core and its relation to the hydrophobic effect. Protein Science: A Publication of the Protein Society 7:158.

68. Gassner NC, Baase WA, Matthews BW (1996) A test of the" jigsaw puzzle" model for protein folding by multiple methionine substitutions within the core of T4 lysozyme. Proceedings of the National Academy of Sciences of the United States of America 93:12155.

69. Pop C, Chen YR, Smith B, Bose K, Bobay B, Tripathy A, Franzen S, Clark AC (2001) Removal of the pro-domain does not affect the conformation of the procaspase-3 dimer. Biochemistry 40:14224-14235.

70. Shiozaki EN, Chai J, Rigotti DJ, Riedl SJ, Li P, Srinivasula SM, Alnemri ES, Fairman R, Shi Y (2003) Mechanism of XIAP-mediated inhibition of caspase-9. Mol Cell 11:519-527.

71. Piana S, Sulpizi M, Rothlisberger U (2003) Structure-based thermodynamic analysis of caspases reveals key residues for dimerization and activity. Biochemistry 42:8720-8728.

72. Batchelor AH, Piper DE, de la Brousse FC, McKnight SL, Wolberger C (1998) The structure of GABPalpha/beta: an ETS domain- ankyrin repeat heterodimer bound to DNA. Science 279:1037-1041.

73. Otwinowski Z, Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. Methods in enzymology 276:307-326.

74. Dodson EJ, Winn M, Ralph A (1997) Collaborative computational Project, number 4: providing programs for protein crystallography. Methods in enzymology:620-633.

75. Jones TA, Zou JY, Cowan S, Kjeldgaard M (1991) Improved methods for building protein models in electron density maps and the location of errors in these models. Acta Crystallographica Section A: Foundations of Crystallography 47:110-119.

76. Demeler B (2005) UltraScan: a comprehensive data analysis software package for analytical ultracentrifugation experiments. Analytical Ultracentrifugation: Techniques And Methods:210-229.

77. Schuck P (2000) Size-distribution analysis of macromolecules by sedimentation velocity ultracentrifugation and lamm equation modeling. Biophysical journal 78:1606-1619.

78. Brown PH (2009) SEDPHAT-An Analysis Platform for the Biophysical Analysis of Reversibly Assembled Multi-protein Complexes in Solution. Biophysical journal 96:74.

79. DeLano WL (2002) The PyMOL Molecular Graphics System DeLano Scientific, San Carlos, CA, USA. http://www.pymol.org.

80. Jackson MR, Laurie TRA (2004) Pocket-Finder Pocket Detection.

81. Hardy JA, Wells JA (2004) Searching for new allosteric sites in enzymes. Curr Opin Struct Biol 14:706-715.

82. Hardy JA, Wells JA (2009) Dissecting an allosteric switch in caspase-7 using chemical and mutational probes. J Biol Chem 284:26063-26069.

83. Russo A, DeGraff W, Friedman N, Mitchell JB (1986) Selective modulation of glutathione levels in human normal versus tumor cells and subsequent differential response to chemotherapy drugs. Cancer Res 46:2845-2848.

84. Jones DP, Brown LAS, Sternberg P (1995) Variability in glutathione-dependent detoxification in vivo and its relevance to detoxication of chemical mixtures. Toxicology 105:267-274.

85. Putt KS, Chen GW, Pearson JM, Sandhorst JS, Hoagland MS, Kwon JT, Hwang SK, Jin H, Churchwell MI, Cho MH, Doerge DR, Helferich WG, Hergenrother PJ (2006) Small-molecule activation of procaspase-3 to caspase-3 as a personalized anticancer strategy. Nat Chem Biol 2:543-550.

86. Denault JB, Drag M, Salvesen GS, Alves J, Heidt AB, Deveraux Q, Harris JL (2007) Small molecules not direct activators of caspases. Nature Chemical Biology 3:519-519.

87. Peterson QP, Goode DR, West DC, Ramsey KN, Lee JJ, Hergenrother PJ (2009) PAC-1 activates procaspase-3 in vitro through relief of zinc-mediated inhibition. J Mol Biol 388:144-158.

88. Witkowski WA, Hardy JA (2009) L2' loop is critical for caspase-7 active site formation. Protein Sci 18:1459-1468.

89. Sowdhamini R, Srinivasan N, Shoichet B, Santi DV, Ramakrishnan C, Balaram P (1989) Stereochemical modeling of disulfide bridges. Criteria for introduction into proteins by site-directed mutagenesis. Protein Engineering Design and Selection 3:95.

90. Boucher D, Blais V, Drag M, Denault JB (2010) Molecular determinants involved in activation of caspase 7. Biosci Rep epub ahead of print.

91. Lesort M, Lee M, Tucholski J, Johnson GV (2003) Cystamine inhibits caspase activity. Implications for the treatment of polyglutamine disorders. J Biol Chem 278:3825-3830.

92. Weik M, Ravelli RB, Kryger G, McSweeney S, Raves ML, Harel M, Gros P, Silman I, Kroon J, Sussman JL (2000) Specific chemical and structural damage to proteins produced by synchrotron radiation. Proc Natl Acad Sci U S A 97:623-628.

93. Burmeister WP (2000) Structural changes in a cryo-cooled protein crystal owing to radiation damage. Acta Crystallographica Section D: Biological Crystallography 56:328-341.

94. Romanowski MJ, Scheer JM, O'Brien T, McDowell RS (2004) Crystal structures of a ligand-free and malonate-bound human caspase-1: implications for the mechanism of substrate binding. Structure (Camb) 12:1361-1371.

95. Podobnik M, Kuhelj R, Turk V, Turk D (1997) Crystal structure of the wild-type human procathepsin B at 2.5 ≈ resolution reveals the native active site of a papain-like cysteine protease zymogen1. Journal of molecular biology 271:774-788.

96. Strajbl M, Florian J, Warshel A (2001) Ab initio evaluation of the free energy surfaces for the general base/acid catalyzed thiolysis of formamide and the hydrolysis of methyl thiolformate: a reference solution reaction for studies of cysteine proteases. The Journal of Physical Chemistry B 105:4471-4484.

97. Ma K, Temiakov D, Anikin M, McAllister WT (2005) Probing conformational changes in T7 RNA polymerase during initiation and termination by using engineered disulfide linkages. Proceedings of the National Academy of Sciences of the United States of America 102:17612.

98. Seeger MA, Von Ballmoos C, Eicher T, Brandstotter L, Verrey F, Diederichs K, Pos KM (2008) Engineered disulfide bonds support the functional rotation mechanism of multidrug efflux pump AcrB. Nature structural & molecular biology 15:199-205.

99. Mitchinson C, Wells JA (1989) Protein engineering of disulfide bonds in subtilisin BPN'. Biochemistry 28:4807-4815.

100. Jeong MY, Kim S, Yun CW, Choi YJ, Cho SG (2007) Engineering a de novo internal disulfide bridge to improve the thermal stability of xylanase from Bacillus stearothermophilus No. 236. Journal of biotechnology 127:300-309.

101. Falcon CM, Swint-Kruse L, Matthews KS (1997) Designed disulfide between N-terminal domains of lactose repressor disrupts allosteric linkage. Journal of Biological Chemistry 272:26818.

102. Qureshi SH, Yang L, Manithody C, Iakhiaev AV, Rezaie AR (2009) Mutagenesis Studies toward Understanding Allostery in Thrombin. Biochemistry 48:8261-8270.

103. Shandiz AT, Capraro BR, Sosnick TR (2007) Intramolecular cross-linking evaluated as a structural probe of the protein folding transition state. Biochemistry 46:13711-13719.

104. Lee JK, Prussia A, Snyder JP, Plemper RK (2007) Reversible inhibition of the fusion activity of measles virus F protein by an engineered intersubunit disulfide bridge. Journal of virology 81:8821.

105. Muslin E, Li D, Stevens F, Donnelly M, Schiffer M, Anderson L (1995) Engineering a domain-locking disulfide into a bacterial malate dehydrogenase produces a redox-sensitive enzyme. Biophysical journal 68:2218-2223.

106. Shimaoka M, Lu C, Palframan RT, Von Andrian UH, McCormack A, Takagi J, Springer TA (2001) Reversibly locking a protein fold in an active conformation with a disulfide bond: integrin L I domains with high affinity and antagonist activity in vivo. Proceedings of the National Academy of Sciences of the United States of America 98:6009.

107. Tomishige M, Vale RD (2000) Controlling kinesin by reversible disulfide cross-linking. Identifying the motility-producing conformational change. J Cell Biol 151:1081-1092.

108. Zheng M, Aslund F, Storz G (1998) Activation of the OxyR transcription factor by reversible disulfide bond formation. Science 279:1718.

109. Singh R, Lillard JW, Jr. (2009) Nanoparticle-based targeted drug delivery. Exp Mol Pathol 86:215-223.

110. Zhou Q, Snipas S, Orth K, Muzio M, Dixit VM, Salvesen GS (1997) Target protease specificity of the viral serpin CrmA. Analysis of five caspases. J Biol Chem 272:7797-7800.

111. Stennicke HR, Salvesen GS (1999) Caspases: preparation and characterization. Methods 17:313-319.

112. Cleland W (1964) Dithiothreitol, a New Protective Reagent for SH Groups*. Biochemistry 3:480-482.

113. Millis KK, Weaver KH, Rabenstein DL (1993) Oxidation/reduction potential of glutathione. The Journal of Organic Chemistry 58:4144-4146.

114. Kabsch W (2010) Xds. Acta Crystallogr D Biol Crystallogr 66:125-132.

115. McCoy AJ, Grosse-Kunstleve RW, Adams PD, Winn MD, Storoni LC, Read RJ (2007) Phaser crystallographic software. J Appl Crystallogr 40:658-674.

116. Emsley P, Lohkamp B, Scott WG, Cowtan K (2010) Features and development of Coot. Acta Crystallogr D Biol Crystallogr 66:486-501.

117. Murshudov GN, Vagin AA, Dodson EJ (1997) Refinement of macromolecular structures by the maximum-likelihood method. Acta Crystallogr D Biol Crystallogr 53:240-255.

118. Pokala N, Handel TM (2005) Energy functions for protein design: adjustment with protein-protein complex affinities, models for the unfolded state, and negative design of solubility and specificity. J Mol Biol 347:203-227.

119. Hellinga HW, Richards FM (1991) Construction of new ligand binding sites in proteins of known structure* 1:: I. Computer-aided modeling of sites with pre-defined geometry. Journal of molecular biology 222:763-785.

120. Chowdry AB, Reynolds KA, Hanes MS, Voorhies M, Pokala N, Handel TM (2007) An object oriented library for computational protein design. Journal of Computational Chemistry 28:2378-2388.

121. Frisch M, Trucks G, Schlegel H, Scuseria G, Robb M, Cheeseman J, Montgomery Jr J, Vreven T, Kudin K, Burant J (2004) Gaussian 03, revision c. 02, Gaussian. Inc, Wallingford, CT 4.

122. Jorgensen WL, Tirado-Rives J (1988) The OPLS [optimized potentials for liquid simulations] potential functions for proteins, energy minimizations for crystals of cyclic peptides and crambin. Journal of the American Chemical Society 110:1657-1666.

123. Laskowski RA (1995) SURFNET: a program for visualizing molecular surfaces, cavities, and intermolecular interactions. Journal of molecular graphics 13:323-330.

124. Foster IT. 1995. Designing and building parallel programs: concepts and tools for parallel software engineering, Addison-Wesley.

125. Simon SR, Cantor CR (1969) Measurement of ligand-induced conformational changes in hemoglobin by circular dichroism. Proceedings of the National Academy of Sciences of the United States of America 63:205.

126. Greenwald J, Le V, Butler SL, Bushman FD, Choe S (1999) The mobility of an HIV-1 integrase active site loop is correlated with catalytic activity. Biochemistry 38:8892-8898.

127. Carr CM, Kim PS (1993) A spring-loaded mechanism for the conformational change of influenza hemagglutinin. Cell 73:823-832.

128. Petsko GA, Ringe D. 2004. Protein structure and function, Sinauer Associates Inc.

129. Miller M, Schneider J, Sathyanarayana BK, Toth MV, Marshall GR, Clawson L, Selk L, Kent S, Wlodawer A (1989) Structure of complex of synthetic HIV-1 protease with a substrate-based inhibitor at 2.3 A resolution. Science 246:1149.

130. Weber IT, Miller M, Jaskolski M, Leis J, Skalka AM, Wlodawer A (1989) Molecular modeling of the HIV-1 protease and its substrate binding site. Science 243:928.

131. Matulis D, Kranz JK, Salemme FR, Todd MJ (2005) Thermodynamic stability of carbonic anhydrase: measurements of binding affinity and stoichiometry using ThermoFluor. Biochemistry 44:5258-5266.

132. Gonzalez M, Argarana CE, Fidelio GD (1999) Extremely high thermal stability of streptavidin and avidin upon biotin binding. Biomolecular engineering 16:67-72.

133. Turingan RS. From an initially transcribing to a processively elongating T7 RNA polymerase: Fluorescence resonance energy transfer (FRET) tests of structural models. (2007). University of Massachusetts Amherst.

134. Matsumura M, Matthews BW (1989) Control of enzyme activity by an engineered disulfide bond. Science 243:792.

135. Ghosh P, Han G, De M, Kim CK, Rotello VM (2008) Gold nanoparticles in delivery applications. Advanced drug delivery reviews 60:1307-1315.

136. Brasnjevic I, Steinbusch HWM, Schmitz C, Martinez-Martinez P (2009) Delivery of peptide and protein drugs over the blood-brain barrier. Progress in neurobiology 87:212-251.

137. Ghosh P, Yang X, Arvizo R, Zhu ZJ, Agasti SS, Mo Z, Rotello VM (2010) Intracellular Delivery of a Membrane-Impermeable Enzyme in Active Form Using Functionalized Gold Nanoparticles. J Am Chem Soc 132:2642-2645.

138. Thornberry NA, Rano TA, Peterson EP, Rasper DM, Timkey T, Garcia-Calvo M, Houtzager VM, Nordstrom PA, Roy S, Vaillancourt JP (1997) A combinatorial approach defines specificities of members of the caspase family and granzyme B. Journal of Biological Chemistry 272:17907.

139. Schilling O, Overall CM (2008) Proteome-derived, database-searchable peptide libraries for identifying protease cleavage sites. Nature biotechnology 26:685-694.

140. Demon D, Van Damme P, Berghe TV, Deceuninck A, Van Durme J, Verspurten J, Helsens K, Impens F, Wejda M, Schymkowitz J (2009) Proteome-wide substrate analysis indicates substrate exclusion as a mechanism to generate caspase-7 versus caspase-3 specificity. Molecular & Cellular Proteomics 8:2700.